

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 8, 2003, 16:19:57 ; Search time 46.2838 Seconds
(without alignments)
4811.606 Million cell updates/sec

Title: US-10-002-309B-2_COPY_24_886

Perfect score: 4597

Sequence: 1 ADNSAIYFNTSQPINDLQ.....AWNGQYLDfSKPRMVRVYK 863

Scoring table: BLQSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_23:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phase:*

10: sp_plant:*

11: sp_rhodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_rvirus:*

16: sp_bacteriap:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4597	100%	898	082882	082882 escherichia
2	1436	31.2	310	092AL1	Q92AL1 escherichia
3	686.5	14.9	1002	09AGX6	Q9AGX6 vibrio chol
4	671	14.6	1335	09KX18	Q9KX18 vibrio chol
5	571	12.4	875	16 Q8D5P4	Q8D5P4 vibrio vuln
6	290	6.3	957	16 Q9KTH2	Q9KTH2 vibrio chol
7	166.5	3.6	1217	16 Q97GW3	Q97GW3 clostridium
8	161	3.5	1038	2 Q45780	Q45780 bacteroides
9	159.5	3.5	2367	2 Q46034	Q46034 clostridium
10	158.5	3.4	2204	5 Q8IL70	Q8IL70 plasmodium
11	156.3	3.4	1157	16 Q97F62	Q97F62 clostridium
12	154.5	3.4	2367	2 Q9F931	Q9F931 clostridium
13	149	3.2	1284	2 Q8RQ09	Q8RQ09 bacillus gl
14	148.5	3.2	2468	16 Q912M3	Q912M3 pseudomonas
15	148.5	3.2	3692	2 Q8FR83	Q8FR83 fusobacteri
16	148	3.2	1021	2 Q46085	Q46085 clostridium

17	148	3.2	2937	16 Q8D990	Q8D990 vibrio vuln
18	147.5	3.2	2306	16 Q8P107	Q8P107 leptospira
19	147	3.2	1386	16 Q92DL0	Q92DL0 listeria in
20	146.5	3.2	1377	16 Q8QZL8	Q8QZL8 mycoplasma
21	146	3.2	1093	16 Q8EV69	Q8EV69 mycoplasma
22	145.5	3.2	2189	5 Q8IKV6	Q8IKV6 plasmodium
23	143.5	3.1	1395	2 Q9AISO	Q9AISO staphylococ
24	143	3.1	1284	16 Q8EV71	Q8EV71 mycoplasma
25	141.5	3.1	2057	2 Q9RE05	Q9RE05 leuconostoc
26	140	3.0	1135	1 Q9P9D1	Q9P9D1 uncultured
27	140	3.0	1196	17 Q8TS48	Q8TS48 methanosarc
28	139	3.0	1110	17 Q8PXT0	Q8PXT0 methanosarc
29	139	3.0	1530	16 Q8Y479	Q8Y479 listeria mo
30	138.5	3.0	1270	16 Q8XAN9	Q8XAN9 escherichia
31	138.5	3.0	1575	2 Q9LCH3	Q9LCH3 streptococ
32	138	3.0	958	17 Q8TLL4	Q8TLL4 methanosarc
33	138	3.0	2894	17 Q58791	Q58791 methanococ
34	137	3.0	5559	16 Q8ZKG6	Q8ZKG6 salmonella
35	136.5	3.0	1457	2 Q9F289	Q9F289 yersinia pe
36	136.5	3.0	3705	2 Q9F285	Q9F285 yersinia pe
37	136.5	3.0	3705	16 Q8ZHA1	Q8ZHA1 yersinia pe
38	136.5	3.0	3710	16 Q8CZU2	Q8CZU2 yersinia pe
39	136	3.0	1677	16 Q8RGN7	Q8RGN7 fusobacteri
40	136	3.0	2167	16 Q92EK5	Q92EK5 listeria in
41	135.5	2.9	2529	16 Q25579	Q25579 helicobacte
42	135	2.9	1700	16 Q8XB95	Q8XB95 escherichia
43	135	2.9	1927	2 Q54875	Q54875 streptococ
44	135	2.9	9904	16 Q8NMQ6	Q8NMQ6 staphylococ
45	134.5	2.9	4727	2 Q8GR92	Q8GR92 mycoplasma

ALIGNMENTS

RESULT 1

ID 082882 PRELIMINARY; PRT; 898 AA.

AC 082882; Q92G01;

DT 01-NOV-1998 (TrEMBLrel. 08, Created)

DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)

DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)

DE Plasmid pO157 DNA, complete sequence.

GN L7031.

OS Escherichia coli O157:H7.

OC Plasmid pO157.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Escherichia.

OX NCBI_TaxID=83334;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=O157:H7;

RX MEDLINE=98290540; PubMed=9628576;

RA Makino K., Ishii K., Yasunaga T., Hattori M., Yokoyama K.,

RA Yutsudo H.C., Kubota Y., Yamauchi Y., Iida T., Yamamoto K., Honda T.,

RA Han C.G., Ohtsubo E., Kasamatsu M., Hayashi T., Kuhara S.,

RA Shinagawa H.;

RT *Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an

RT enterohemorrhagic Escherichia coli O157:H7 derived from Sakai

RT outbreak*;

RL DNA Res. 5:1-9(1998).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=EDL933;

RX MEDLINE=98391744; PubMed=9722640;

RA Burland V., Shao Y., Perna N.T., Plunkett G., Sofia H.J.,

RA Blattner F.R.;

RT *The complete DNA sequence and analysis of the large virulence plasmid

RT of Escherichia coli O157:H7*;

RL Nucleic Acids Res. 26:4196-4204(1998).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=EDL933;

RA Bruner W.;

RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.

```

DR EMBL: AB011549; BAA31757.3; -.
DR EMBL: AF074613; AAC70099.1; -.
DR EMBL: Y11831; CAA72517.1; -.
KW Hypothetical protein; Plasmid.
SQ SEQUENCE 898 AA; 95548 MW; 3CIAE23E3EAE1FAB CRC64;

Query Match
Best Local Similarity 100.0%; Score 4597; DB 2; Length 898;
Matches 863; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ADNSAIYFNTSOPINDLOGSLAAEYKFAQSOLTPAHKPEGDSQPHLTSLRKSLLLVRPV 60
Db 36 ADNSAIYFNTSOPINDLOGSLAAEYKFAQSOLTPAHKPEGDSQPHLTSLRKSLLLVRPV 95

Qy 61 KADDKTPVQVEARDNNKILGTTLTPSSLPDTIYHLDGVPGEIDFTPHNGTKKIINT 120
Db 96 KADDKTPVQVEARDNNKILGTTLTPSSLPDTIYHLDGVPGEIDFTPHNGTKKIINT 155

Qy 121 VAEVKNLSASGSSISHLTNNALVEIHTANGRWVRDIYLPQDLEGKMWRFVSSAGYS 180
Db 156 VAEVKNLSASGSSISHLTNNALVEIHTANGRWVRDIYLPQDLEGKMWRFVSSAGYS 215

Qy 181 STVFYGDRTVLSVGNLTLLFKYVNGQWFRSGELENRRITYAQHINSAELPAHWIYVPLNL 240
Db 216 STVFYGDRTVLSVGNLTLLFKYVNGQWFRSGELENRRITYAQHINSAELPAHWIYVPLNL 275

Qy 241 VIKQGNLSGRINDIKIGAPGELLHTIDIGMLTTP 275
Db 276 VIKQGNLSGRINDIKIGAPGELLHTIDIGMLTTP 310

Qy 301 IYNNYAPLHLKEVMLPTGELLTMDPGNGWHSQTMQRIGKELVSHGIDNANYGLNSTA 360
Db 336 IYNNYAPLHLKEVMLPTGELLTMDPGNGWHSQTMQRIGKELVSHGIDNANYGLNSTA 395

Qy 361 GLGENSHPYVQAIAHNSRGNYANGIQVHGSGGGGIVTLDSITGNFSEVHGNYGLG 420
Db 396 GLGENSHPYVQAIAHNSRGNYANGIQVHGSGGGGIVTLDSITGNFSEVHGNYGLG 455

Qy 421 HYVDGFGSVHRSAAENNTSGWDGDKRFIPNFYPSQTNESKSLNNOCEPFDGKFGF 480
Db 456 HYVDGFGSVHRSAAENNTSGWDGDKRFIPNFYPSQTNESKSLNNOCEPFDGKFGF 515

Qy 481 DAMAGGSPFAANRFTMYTPNSAIIQRFENKAVFDSRSSTGFSKWNADTOEMEPYEH 540
Db 516 DAMAGGSPFAANRFTMYTPNSAIIQRFENKAVFDSRSSTGFSKWNADTOEMEPYEH 575

Qy 541 IDRAEQITASVNELSKMAELMAEYAVVYVHWMNGNWTNLIYIPTASADNRGSLTINH 600
Db 576 IDRAEQITASVNELSKMAELMAEYAVVYVHWMNGNWTNLIYIPTASADNRGSLTINH 635

Qy 601 EAGYNSYLFINGDEKVVYSGYKSFYSDGQFKERDVVDVTREARKPEQGVPTTLVGY 660
Db 636 EAGYNSYLFINGDEKVVYSGYKSFYSDGQFKERDVVDVTREARKPEQGVPTTLVGY 695

Qy 661 DPEGTLSYIYPAMYGYGFTYSDSQNSDNDQLOVDTKEGQLRFLANHRANNTVMN 720
Db 696 DPEGTLSYIYPAMYGYGFTYSDSQNSDNDQLOVDTKEGQLRFLANHRANNTVMN 755

Qy 721 KFHINVPESQPTQATLVCKNKLIDTKSLTPAPEGLTYTVNGQALPAKENEGCIVSVNSG 780
Db 756 KFHINVPESQPTQATLVCKNKLIDTKSLTPAPEGLTYTVNGQALPAKENEGCIVSVNSG 815

Qy 781 KRYCLPVGQSGYSLPDMIVGQEVYVDSGAKAKVLLSDWNLISYNRIGEFVGNVPADMK 840
Db 816 KRYCLPVGQSGYSLPDMIVGQEVYVDSGAKAKVLLSDWNLISYNRIGEFVGNVPADMK 875

Qy 841 KYAAWNGQYLDKSPSRMRVYK 863
Db 876 KYAAWNGQYLDKSPSRMRVYK 898

RESULT 2

```

```

Q9ZAL1
ID Q9ZAL1 PRELIMINARY; PRT; 310 AA.
AC Q9ZAL1;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Hypothetical 34.0 kDa protein (Fragment).
OS Escherichia coli.
OG Plasmid p0157.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EDL933;
RA Brunder W.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y11275; CAA72142.1; -.
KW Hypothetical protein; Plasmid.
FT NON_TER 310 310
SQ SEQUENCE 310 AA; 34032 MW; 2AA72C166F315BA6 CRC64;

Query Match
Best Local Similarity 100.0%; Score 1436; DB 2; Length 310;
Matches 275; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ADNSAIYFNTSOPINDLOGSLAAEYKFAQSOLTPAHKPEGDSQPHLTSLRKSLLLVRPV 60
Db 36 ADNSAIYFNTSOPINDLOGSLAAEYKFAQSOLTPAHKPEGDSQPHLTSLRKSLLLVRPV 95

Qy 61 KADDKTPVQVEARDNNKILGTTLTPSSLPDTIYHLDGVPGEIDFTPHNGTKKIINT 120
Db 96 KADDKTPVQVEARDNNKILGTTLTPSSLPDTIYHLDGVPGEIDFTPHNGTKKIINT 155

Qy 121 VAEVKNLSASGSSISHLTNNALVEIHTANGRWVRDIYLPQDLEGKMWRFVSSAGYS 180
Db 156 VAEVKNLSASGSSISHLTNNALVEIHTANGRWVRDIYLPQDLEGKMWRFVSSAGYS 215

Qy 181 STVFYGDRTVLSVGNLTLLFKYVNGQWFRSGELENRRITYAQHINSAELPAHWIYVPLNL 240
Db 216 STVFYGDRTVLSVGNLTLLFKYVNGQWFRSGELENRRITYAQHINSAELPAHWIYVPLNL 275

Qy 241 VIKQGNLSGRINDIKIGAPGELLHTIDIGMLTTP 275
Db 276 VIKQGNLSGRINDIKIGAPGELLHTIDIGMLTTP 310

RESULT 3
Q9AGX6
ID Q9AGX6 PRELIMINARY; PRT; 1002 AA.
AC Q9AGX6;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE ToxR-activated gene A protein.
CN TAGA.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=395;
RA MEDLINE=21117025; PubMed=11179381;
RA Karalis D.K.R., Lan R., Kaper J.B., Reeves P.R.;
RT "Comparison of Vibrio cholerae Pathogenicity Islands in Sixth and
RT Seventh Pandemic Strains";
RL Infect. Immun. 69:1947-1952(2001).
CC -1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
DR EMBL; AF325733; AAK20748.1; -.
DR InterPro; IPR003961; FN_III.
DR Pfam; PF00041; fn3; 1.
DR SMART; SM00060; FN3; 1.

```

kw	Hydrolase.	1002 AA; 116465 MW; 6A239DFB6408407F CRC64;
SQ	SEQUENCE	
	Query Match	14.9%; Score 686.5; DB 2: Length 1002;
	Best Local Similarity	24.08; Pred No. 1.9e-36;
	Matches 230; Conservative	157; Mismatches 275; Indels 295; Gaps 40;
QY	4 NSAIYNTSOPINDLOGSLAAEKFAQSOLPAHPKKGDSQPHLTSLRKSLLLRP-VKA	62
DB	137 NEIDYVHTTEIKNVTGSGREVFYTHVI--SPEGRKNNEPIITGRDALILFKPSIKN	194
QY	63 DDKTPVOVEARDONNKILGLTLTYPPSSLDPTIYHLDGPCEGGIDFPHGCTKKLIINTVA	122
DB	195 SSSILMKIYSDEG--LTSVKMKSPLPKTDQPID-----ID-----	230
QY	123 EVNKLSDAGSSIHSLTNALVEIHTANGRWVRDYLPGQPDLEGMRYRVSSAGYST	182
DB	231 -----	230
QY	183 VFTGDRKVTLVGNTLFLKYVGNQFRSGELENNR-ITYAQHIINSAELPAHWIVPGLMLV	241
DB	231 -----ENNVSYSNYSYSAETIPWNRKMSGLH	259
QY	242 IKOGNLSGRL-----NDIKIGAPCELLHTIDICMLTTPDRDFDAKDKEAHR-EYFOTI	295
DB	260 FEDEN--GNLGITESERIKSPASELIIQNDJGLMYKPRGNIVIKELETAVDYFOKV	317
QY	296 PVSRMIYNNTYAPLHKVMPLPTGSLLTDMDPNGGWHSGMTQRIGKELVSHGIDNANYG	355
DB	318 PVSCLIFSVDYTPHFEXKITLPNGSVYTEKSDIGGWGHQMREAVGKALYSTGINNANLG	377
QY	356 LNSTAGLCGENSHPVYAQAHAHNRSRGNYANGIOVHGGSGGGIVLTDLSTGLNESHSHEVCH	415
DB	378 IVASSGYSQ-QYNRLTRHITAHTNIGYVINGVVHGGSGGGIVTLETLHNEMSHLEIGH	436
QY	416 NYGLGHVYDGFKSGVHRSNAENNSTWGDDKRFPNFYPSOIRNEKSCL--NNOCQEPF	473
DB	437 NYGLGHVYAG--GTSH---CPDTISGWGDYYKRFIAFNFWKRSPQSIIRDNQEVWAPF	490
QY	474 -DGHKFQFDAMAGSPF--SAANRFTMTYTPNSSALIIRFFENKAYFSRSTSGFSKNAD	530
DB	491 MDKVTYLDWAMSGCYDHQNGIISRVTLLHPHYARIODWLKNGAVV--INNDEVWMD-E	546
QY	531 TOEMEPYEHTIDRAEQITASVNELSESMAELMAEYAVVYHHMHGNCNTRNIYIPTASAD	590
DB	547 LKNIYVYKGT-----NFKVPYIKGVFVVTILGYVDPDKIN-----PSOLIYPPTY-	591
QY	591 NRGSIL-----TINHEAGYNSY--LFINGDEKVVYSQ-----	619
DB	592 NYGNIIFDLKPRESSELLKGQYVKDVYLDVRNTHWHTMLVNRKEEKICRSYLSPKGCK	651
QY	620 ----GYKK-----SFVSDGQFWKERDVEDVTREARKPEQFGVPTVLVG----Y	659
DB	652 FEFGLGYEDIENKICTGSRSHIYLEDG--KKNP-----ESKYNDYF--LLSIDGGEITS	701
QY	660 YDPGTL-----SSYIYPMYGA-----YGFTYSDSQNL-----	690
DB	702 YVPDSTIGESICKLSMGVTVGGFIKNGSCRODGVFMNGFQWATLNOSGVNSTVTYM	761
QY	691 DNDCOLOVDTKEGQLR-FRLAHRRANNVTMMKFHINVPTSQOPTQATLVCNKKILDTKSL	749
DB	762 SNECVLKIKDNKNITESISIPNYRIEKQNOKIHLNISREKPIIDIINYCGHELTSIKV	821
QY	750 TPAPEGLYTTVNGQALPARENDGCITVSNSGKRYCLPVQRSQSYSLPDMIVGOEYVVDG	809
DB	822 SDNPD-----IKLLKGPDI-----VQOEHGYT-----SYEPKLPSP	852
QY	810 AKAKVLLSDWMLS-YNRIGETFVGWNVPADMKKVKAWNGOYL---DFS-KPRSRRVV	861
DB	853 -----WFKHYNFEPEKNEINHELGMKRVND-----NDEYICRNFNSDSDEMFKV	897

O9KN18

ID		O9KN18	PRELIMINARY;	PRT; 1335 AA.
AC		Q9KN18;		
DT	01-OCT-2000	(TreeBLRel. 15, Created)		
DT	01-OCT-2000	(TreeBLRel. 15, Last sequence update)		
DT	01-DEC-2001	(TreeBLRel. 15, Last annotation update)		
DE		TagA-related protein.		
GN	VCA0148.			
OS	Vibrio cholerae.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;			
OC	Vibrionaceae; Vibrio.			
OX	NCBTaxID=666;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=El Tor N16961 / Serotype O1;			
RX	MEDLINE=20406833; PubMed=10952301;			
RA	Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,			
RA	Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,			
RA	Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,			
RA	Ermolaeva V.M., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,			
RA	McDonald L., Utterback T., Fleischmann R.D., Niemeyer W.C., White O.,			
RA	Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,			
RA	Frazer C.M.;			
RT	"DNA sequence of both chromosomes of the cholera pathogen Vibrio			
RT	cholerae.";			
RL	Nature 406:477-483(2000).			
DR	EMBL; AF004356; AAF96061.1; -.			
DR	TIGR; VCAO148; --			
KW	Complete proteome.			
SQL	SEQUENCE 1335 AA; 149521 MW; 0E339F0DBCBD2DEFD CRC64;			

Query Match 14.6%; Score 671; DB 16; Length 1335;
Best Local Similarity 25.5%; Pred.No. 3.2e+35;
Matches 231; Conservative 108; Mismatches 248; Indels 318; Gaps 36;

QY	16	NDLOQSAAEVKFAQSOLPAHPK--EGDSOPHTLSL--PKSLLLRVPVKADDKTPVQVE	71
DB	353	NHLOSLEGSLISOTHTSVAP-KPGNELTQGHLDAITMNEALLFTPOGGEINQVRAE	411
QY	72	ARDNNKKILGTLTLPSPSSLDPTIYLHDGVPEGGIDFTPHNGTKKIINTVAEYNKLSDAS	131
DB	412	VFLDELVTTLML-PPSAL-----AADQPENGKMKV-----	443
QY	132	GSSHLSHLTNNAIVEIHTANGRWRIYLPQGPDLKGKVRFVSAGYSTVFYGDKRYT	191
DB	444	--VESHLA-----	449
QY	192	LSVGNTLLFKYYNQWFPSGLSENNRRITYAQHIWSAELPAHNIVPGLNLVI----	246
DB	450	-----WSLPLQDMWKPKGLSRULTDNLGREGV	476
QY	247	LGRINDIKIGAPGELLHTIIDIGMTTPRDREFAKD-KEAHREYFQTIPVSRMITVNKY	305
DB	477	LS-QGETFGCAGPELVTONDIGCLPRDRNTMIONTLATDYFOKITASKLYVDAY	534
QY	306	APLHKLEVMPTCELLTDMDPCNGSWHSGMTMRQRIGKEYSHGDINANTGLNSTAGLGEN	365
DB	535	TPAHFPVTMPNGVVYTDKSASTGWHSBGDMREAIGAARVSTGINNAWGIVSSAGYSQQ	594
QY	366	----SHPVYAQLAANRSRGNYAN-----GIOVHGSGGGGIIVTDSLTCHEFSHEVG	414
DB	595	YNRFRNH-----ITAHTNVGYTTCKDDLPQVYVHHGGGGGITVLATGTNEWSEHEL	648
QY	415	HNYGLGHVDFGFSGVHRSAENNSHWCDGXKRFFPNF-----YPSTNKESKLNNQ	468
DB	649	HNYGLGHV---PYNASIH-----DLESGMWDADFHQIFIGNLHWKDGYVTOOGGD-	696
QY	469	CQEPPF-DGKHCFGFDAMAGSP--FSAAARRFTMYTPNSSAIIOEFFEKAVDFSRSSTGS	525
DB	697	IYVPFFKAQFRLDAQNGEQETVGTISRFTLEHPAQSKRAWRMNNGFNLDSESFGTV	756
QY	526	KWNADTQEMEPTYERTIDRAEQTI-----TASNVELSES	557

RESULT 4

229 DVTQOQGD-----DIVPPFKDAFRFLVDAQNGGQEQEYVGTISRFTLEHPAQRKAQRW 288
 510 FENKAVFDSRSSTGFSKWNADTQEMEPYEHTIDRAEQITASVNEL-----SESKMA 560
 283 MNGFNLDSSSPSGYVQWQAAQRYQTVETDTPKPQQTGVAVWTLGLGIYDPYNENPSQI- 341
 561 ELMAEYAVVVMWN-----GNWTRNIYPTA-----SADNRGSLTITINHEAGY 604
 342 -----YPLVTSNYGNVNFELPQGVQAGPQEGWOPVADLTPAELESDSWQTLRMDGEQQRV 396
 605 NSYLF--INGDEKVVSGYKKSFSVSDQFWRKRDVVDTRARKPEQGV--PVTTLVGY 660
 397 CKFTFOAANGDSAVFVGVDQS-----TDRCSGRLQWHINSHNMTSAOGDY 443
 661 DPEGTLSSYIYPAH-----YGAYGFTYSDDSQNLS----- 690
 444 E--LLSKYGRGAVTYTPTPEVGEVTLTLNKSGTDHDGAGFFVVGNNCEQISGYMHKHK 500
 691 -----DNDQQLQVDTKEG---QLRFRLANHRANVTVMKFNHINVPTE 720
 501 TWRYALRGDEVLRPSYQTOGQCQLDVFANGASERVVLNASHRVNDS--NKFHWNLAMD 558
 730 S-OPTQATLVCNKKILDTKSLTPAPEGLTYTVNGQALPAKENEGCIYVNSGKRYCLPVG 786
 559 NGVPTQVRLSCSDRNGETELTRFTPE-----ONPPIDALKGPII-----IG 599
 789 QRSQYS 794
 600 QEQYS 605
 RESULT 6
 ID Q9KTH2 PRELIMINARY; PRT; 957 AA.
 AC Q9KTH2;
 DT 01-OCT-2000 (TRENBIrel. 15, Created)
 DT 01-OCT-2000 (TRENBIrel. 15, Last sequence update)
 DT 01-MAR-2003 (TRENBIrel. 23, Last annotation update)
 DE Hemolysin-related protein.
 GN VC0930.
 OS Vibrio cholerae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrio.
 OC NCBI_TaxID=666;
 [1]
 SEQUENCE FROM N.A.
 RC STRAIN=El Tor N16961 / Serotype O1;
 RX MEDLINE=204008833; PubMed=10952301;
 RA Heidelberg J.F., Eelsen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
 Dodson R.J., Hatt D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
 Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
 Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
 McDonald L., Ueberback T., Fleischmann R.D., Nierman W.C., White O.,
 Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
 RA Fraser C.N.;
 RA "DNA sequence of both chromosomes of the cholera pathogen Vibrio
 cholerae";
 RL Nature 406:477-483(2000).
 DR EMBL; AE0041176; AAF94092.1; --
 DR TIGR; VC0930; --
 DR InterPro; IPR002048; EF-hand.
 DR PROSITE; PS00018; EF_HAND; 1.
 KW Complete proteome.
 SQ SEQUENCE 957 AA; 103989 MW; 4C6BA0FBC755CDAD CRC64;
 Query Match 6.3%; Score 290; DB 16; Length 957;
 Best Local Similarity 53.8%; Pred. No. 2.4e-10;
 Matches 57; Conservative 14; Mismatches 35; Indels 0; Gaps 0
 Qy 756 LTYTVNGQALPAKENEGCIYVNSGKRYCLPVGQSRGYSLPDWIVQEVYDVGAKAVL 815
 Db 13 LSLSSNVVQATNTAEAGCIIIRLNGEKYCLKVGSRGYSLPWSWIYHPVDVQAPSGVSM 72


```

QY      816 LSDWDNLNLYNRIGFVGNGVNPADAKKVKVAWNCQYLDFSKPSMRVV 861
       |||||: | :| |||:| |||||:|
Db      73 LSDWDNLNLYNRILAVFDRTYTGNEDLNKKAYNGAYLDFSKPSMRYL 118

RESULT 7
ID Q97GW3 PRELIMINARY; PRT; 1217 AA.
IC AC O97GW3;
DT 01-OCT-2001 (TReMBLrel. 18, Created)
DT 01-OCT-2001 (TReMBLrel. 18, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Alpha-glucosidase fused to unknown alpha-amylase C-terminal
DE domain.
DE GNA
DN CAC2252.
OS Clostridium acetobutylicum.
OC Bacteridia; Firmicutes; Clostridia; Clostridiales; Clostridia
OC Clostridium.
ON NCBI_TaxId=1488;
OX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Neolling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng
RA Gibson K., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RA "Genome sequence and comparative analysis of the solvent-pro-
RT ducterium Clostridium acetobutylicum.";
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL; AE007725; AAK80209.1; -.
DR InterPro; IPR005085; CBM_25.
DR InterPro; IPR000322; Glyco_hydro_31.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF03423; CBM_25; 1.
DR Pfam; PF01055; Glyco_hydro_31; 1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
KW Complete proteome.
SQ SEQUENCE 1217 AA; 134530 MW; 26F9D66B02688A30 CRC64;
```

Query Match	3.6%	Score 166.5;	DB 16;	Length 1217;
Best Local Similarity	18.6%;	Pred. No. 0.049;		
Matches 179;	Conservative 121;	Mismatches 319;	Indels 343;	Gaps 48;
Qy	56	LVRPVKA-----DDTTPVQVEARDONNKILGTLTLYPPSSLP-----DTIYHLDGVPEGGI	106	
Dd	29	IIRPVSAAAPNMATISIKITVKENARV-----SNLSAKLNGDPTQLIVNGLDETDI	79	
Qy	107	-----DFTPHNGTKKIINTVAEYNKILSDAGSGSIHSLHTNNALV-----EIHTA	150	
Dd	80	KICEPOVLKVDPKP-SGOSSSDTLTVDPNPKNIN-TGNIISDLNSDPWVITQTQMTIKIS	137	
Qy	151	NGRWVRDIYLPQGPDLEGKMV---RFVSSAGYSSTVFYGDRKVTLSVGNTLLFFKYVNGQM	207	
Dd	138	KSDLTMSVY-----DSTGCKIVKQOOSIAKSQSVSFTHNSGR-----FYGINCYN	181	
Qy	208	FR-----SGELEN-NRIITYAQHIWSAELPAHVIVPGNLIVIQGNLSGRND-----	253	
Dd	182	FKEDSNKGMLRNGTESYVQGHCGSPFVWSNDGYGLLVDSGDGSFTTGTSLOYSGIS	241	
Qy	254	-----IKICAPGELLHTIDI-----GMLTTP-----RDREFDAKOEAHRE	290	
Dd	242	KTDTDYILMLGSPKEVISEDSVSKCAPFPKWATGFTNTQGWNNLSLCTGNDEDKLKS	301	
Qy	291	YFQT-----IPVSRMIV-----NNYAPLHLKEVMCP---TGETLTDMDPGNGGWHHS	333	
Dd	302	VLNTVRSKQIPIDNFCLOFNKKMGODNYGEFKWNTDNFPPAQNGOLKAYMD-SKGLKMT	360	
Qy	334	GTMRORI-----KELYSHG-----IDNANTCLNSTAGLGENSEHYVV	370	
Dd	361	GIMPRILIADSEQARYVTSKGWWLPGDSAAASYCSGKKMHENVNFAISDV-----	409	
Qy	371	VAQLAAHNSRNGYANGIOVHGSGGGGGVTITLDSTLGNFEFSHEVCHNYGLGHY-----	422	

Db	410	-----RKWNWNIIO-----DAFDKGIVGTFWDECDENVNTGNGNMMHRA 450
Qy	423	-VDGFGSGYHRSAAENNNSTWG-----WDGDKRFPINFPYSPSTNEK-----462
Db	451	LYDG-----QRAYKNORVSLNRNYAGAQRYCYGMMWGD-----ISTGFDSMANQRMR 500
Qy	463	-SCLNNQOEFPDGHKCFD--AMAGSP-----FSAANRFTMYTPNSSLAIQHFENKAV 515
Db	501	LSAVN-----LGEAKWGMGTGGFNDGDPPTENYARMMEESAFTP-----IFRVHGODNRV 550
Qy	516	-----FDSRSSTGFSGKNADTOEMEPYEHTIDRAEQITA-----SVNELSESK 558
Db	551	RYPWAFGSTAEAAAKAKMLRYTLPIIYYSVDRSASQSGGLYVSLMMEYPNDSNAANDK 610
Qy	559	MAELMAEYAVVYKHMNGNWRNTRYIP-----TASADNRGSLITINHEAGVNSY-----607
Db	611	EAWMEGDYMLVSPVYNQGTQSKSYLPGNNIDYTTGREYTGQ--QTINTAVDSTNNMSDIP 669
Qy	608	LFING-----DEKVVSGYKKSPVSD-----GQFWKKR 635
Db	670	LFKSGAIIPTODFENYGEKKITDYYVDAPFSDKATTFDYDDDGTSYDENGSTFDOR 729
Qy	636	DYVDTREARKPQFGVPVYTLVGYDPEGLTSSYI-----YP 672
Db	730	MTLQTSSTDSKSVQFNIDKNT--GSTYPD--LKDYIVKMHYKMGCAVVTANGOALTOYSSYD 785
Qy	673	AMYGAYGFTYSDDSQNLSDNCQLQVDTKEQLRFLANHRANTYVANKHINVPTESQP 732
Db	786	ALKSASGEGYASGT-----DTYGNVYIKYSSGDAKN-----INVSCNPLP 826
Qy	733	TOATLVCKNK-----ILDTKSLTPAPEGLT--YTVNGQALPAKENEGCIVSVNSGKRYCL 785
Db	827	VITTAANPKFGTYGYPQTVSLTASKSDATYYIL-----DGTTFVNSTK-YTA 875
Qy	786	PV 787
Db	876	PI 877
RESULT 8		
ID	Q45780	
ID	Q45780	PRELIMINARY; PRT: 1038 AA.
AC	Q45780;	
DT	01-NOV-1996 (TrEMBLrel. 01, Created)	
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)	
DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)	
DE	Outer membrane protein.	
GN	SUSC.	
OS	Bacteroides thetaiotaomicron.	
OC	Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;	
OC	Bacteroidaceae; Bacteroides.	
OX	NCBI_TaxID=818;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=BT55482;	
RC	MEDLINE=96146534; PubMed=8550519;	
RA	Reeves A.R., D'Elia J.N., Frias J., Salyers A.A.;	
RT	"A Bacteroides thetaiotaomicron outer membrane protein that is	
RT	essential for utilization of maltooligosaccharides and starch.";	
RL	J. Bacteriol. 178:823-830(1996).	
CC	- - SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).	
DR	EMBL; L49338; AAA95938.1;	
DR	InterPro; IPR001395; Aldo/ket.red.	
DR	InterPro; IPR000531; TonB_boxC.	
DR	Pfam; PF00593; TonB_dep_Rec; 1.	
DR	PROSITE; PS00063; ALDOKETO_REDUCTASE_3; 1.	
DR	Membrane; outer membrane; Receptor; TonB box.	
SW	SEQUENCE 1038 AA; 115886 MW; DD4C6CE38A2DC5E1 CRC64;	

Query Match 3.5%; Score 161; DB 2; Length 1038;
Best Local Similarity 18.1%; Pred. No. 0.088;
Matches 180; Conservative 120; Mismatches 328; Indels 36

```
QY 16 NDQGSLLAEVKEFAQSILPAHPKEGDSOPHLASRLKSLLLVR-----PVKADDKTP---V 68
DB 131 NDMTGSVMA-----IKDELSKG-----IITNAQDMLSGKIAGVSVISNDGTPOGGA 177
QY 69 QVEARDNNKILGTLTLYPSSLPDIYHLDGVPPEGIDFTPHNGTKKIINTVAENVK--- 126
DB 178 QIRIRGSSSL-----NASNDPLIVDGL---AID---NEGIKGMANGLSMVPAD 221
QY 127 -----LSDASGSSISHLTNNAIVEHTANGRWVRDIYLPQGPDLGKMWRFVSSAGYS 180
DB 222 IETLVTKDASATAYGSRASNGVIITTKKG-----NGQAPSVTYNGSVSFS 270
QY 181 STVYFGDKVTLSVGNLTLLFKYNGQWFRSGELENRRITYAQHIWSAELPA-----HW 233
DB 271 KT-----QKRYDVLSDGEYRA-----YANQLMGDKLPADLGANTDW 307
QY 234 -----IVPGLMLVTQGG--NLSGRINDIKIGAPGELLHLTHIDIGMLTPDRDFPAK 283
DB 308 QDQIFRTAVSTDHHSVINGGFKNLPYRVS-----LGYTDDNGIVKTSNFR-RFTA 356
QY 284 DKEAHREYFQ---TIPYSRMIV---NNYAPLHLKE---VMLPTGELLTMDP---GNGG 330
DB 357 SVNLPAPFFEDHLKFNINAKFMKNGKNRYADSRKYWRALAIIDTPRPYSNEDPQFTGGY 416
QY 331 WHSGTMRQRIKELVSHGIDNANTGLNSTAGLGE-----NSHPYVVAQLAAHNSRG 381
DB 417 WQN-----INSTGFSNPDMKYTSNPNSSPONPLAALKLNKDKG 454
QY 382 NYANGIQVHGGSGGGIVTLDSTLGNFESHEVGHNYGL-----GHYVDGFKG 428
DB 455 N-----SNDYFGVNDVYKFEHFLPDLRLHASIGGEYAEQTOT 491
QY 429 SVHRSANNSTMGWGD---KKRFTPNFVPSOTNEKSKLNNOCQEPFDGKHGFGDAMA 484
DB 492 TIVSPYFEGNNYVGNWGDVQYKYNLSYVQVI--KSLGAND-----FLIMV 538
QY 485 GGSFPFSAANRFTMYTPNSSAIQIRFFENKAVFDSRSSTGFSK---WNAQTOEMPEYHTI 541
DB 539 GGE-----QHFRN-----GPEEGQGWDSYTOE-----PHDA 566
QY 542 DRAGQITASVNELSESKMA---ELMAEYAVVVMHNGWNRNIIYPTASADNRGSILT 597
DB 567 KLREQATAYATRLTVSYFGLRNLNRLRYLFTFTMRWDGS-----SRFSKDNRW--- 615
QY 598 INHEAGYNSYLFINGDEKVVSGYKFSVSDGQFWKERDVVDYTRAEARKPEQFGVPVTLV 657
DB 616 -----GTFPSLALGHK---IKREENFLKDVNVLSDLKLR---LHGWTGQQ 654
QY 658 GYDPEGTLSYIYPAMYGAYGF--TYSDDS--QNLSDNDCOLQVDTKEGOLRFLANHR 713
DB 655 NIGDDFAYLPLYVYVNNYEAQYPPGDTYTSRPRKAFENLKWETTTWNAGLDFGFLNGR 714
QY 714 ANNTVMKFNHNVTESQPTQATLVCNKKILDTSLPAPSGLY-----TVNGQ 763
DB 715 ITGIDGYF-----RKTCTVTLRSPNNILOCPDDEYRFTGKLYRFGFSINAK 763
QY 764 ALPAKE-----NEGIVSVNSG--KRYCLPVG----- 788
DB 764 PIVTKDFTWDLSYNITWNHNEITKLTDGDDSDYVEAGDKLSRGNNTKVQAHKGYAANS 823
QY 789 -----QRSQSYSLPDWIVGQEVYVDSGAKAKVLLSDNLSYNRIGYFVGNNV 835
DB 824 FYVSRGNNTKVQAHKGYAANSFYVYQVYDENGKPI-----ENMFVDRNGN--GTID 874
QY 836 PADMKKVTAWNG-----OYLDFSKPKSMR 859
DB 875 SGDKYIYKKPAGDVLMLGLTSKMQYKNFDFSESLR 908
```

RESULT 9

Q46034

PRELIMINARY: PRT: 2367 AA.

```
Q46034;
AC 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
DE Toxin B.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1496;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-isolate 1470;
RX MEDLINE=96079281; PubMed=7494480;
RA von Eichel-Streiber C., Meyer zu Heringdorf D., Habermann E.,
RA Sartingen S.;
RT *closing in on the toxic domain through analysis of a variant
RT Clostridium difficile cytotoxin B.*;
RL Mol. Microbiol. 17:313-321(1995).
DR EMBL; Z23277; CAAB0815.1;
DR InterPro; IPR002479; CWL-binding.
DR InterPro; IPR001950; TIF_SUI1.
DR Pfam; PF01473; CW-binding_1; 18.
DR PROSITE; PS01118; SUI1_1; 1.
SQ SEQUENCE 2367 AA: 269186 MW: EF9823DAE70427F3 CRC64;

Query Match 3.5%; Score 159.5; DB 2; Length 2367;
Best Local Similarity 18.4%; Pred. No. 0.4;
Matches 185; Conservative 149; Mismatches 321; Indels 353; Gaps 49;

QY 4 NSAIYNTSOPINDLOGSLAAEVKEFAQSILPAHPKEGDSOPHLTSLRKSLLLRVPKAD 63
DB 1454 NSELQKNIPYFVDSGK-----ENGFINGSTKEG-----LFVSELFDVWLISKVYMD 1501
QY 64 DKTP-----VQVEARDNNKILG-----TLTYPESSLPDIYHLDGV 101
DB 1502 DSPSPGYTSNNLKDKVYTKDNVNLITGYLLKDDIKISLSLTLDQEKIKLSVHLD-- 1559
QY 102 PEGGIDFTPHNGTKKIINTVAENVKLSLDASGSIHSLTNNAIVEHTANGRWVRDIYLP 161
DB 1560 -ESG-----VAEILKPMNRKGS---TNTSDSLMSFLESNN---IKSIFV- 1596
QY 162 QGPDLEGKMYRFVSSAGY---SSTVYFG-----DRKVTLSVGN 197
DB 1597 ---NFLQSNKIFILDANFIISGTSIGQFEFICDENNNIQPIKFTNTLTLYVGNR 1653
QY 198 L-LFKYVNGQWFRSGELENRRITYAO-HWSAELPAHWIIVPGLNLVKKQNLGS----- 249
DB 1654 QNWIVPEYDLDSGDISSTVINFPSQYLYGIDSCVKNKVISNIYTDENIITPYETNN 1713
QY 250 -----RLNDIKI-----GAPGELLHTIDIGMLTPDRDF-DFAK 283
DB 1714 TYPEVIVLDANYINEKINVININDLSIRYVNSDNDGNDFILMSTSEENKVSQKIRFVNVFK 1773
QY 284 DKE-AHREYF---QTIPYSRMIVVNYAPLHLKEVNL----- 315
DB 1774 DKTLANLSFNSDKQDVPVSEIIL-SFTPSYEDGLGYDGLGLVSLYNEKFIYNNFGMM 1832
QY 316 -----PTGELLTD-----MDPGNGGWSHSGTMRQRIKELVSHGI 349
DB 1833 VSGLIYINDSLYFKPPVNNLITGFTVVGDKYFYFNPINGGAAS-----IGETI---I 1862
QY 350 DNANYGLNSTAGLGENSHPYVVAQLAAHNSRGNY-----ANGIQVHGGSGGGIITLSD 404
DB 1883 DDKNYFNQSG-----VLQTVGFSTEDGFKYFAPAN-----TLDEN 1918
QY 405 LGNE---FSHEVGHNTYGLGHYVDGFKGSHRSBAENNNSTWGW---DGDKKRFPINFPY 458
DB 1919 LEGEADFTCKLIIDENIYFYFEDNYRGAVE-----WKELDGMHYFSPE----- 1962
QY 459 TNEKSLNNOCQEPFDGKHGKFGFDAMAGGSPFSAANRF--TWYTPNSSAIQIRFF----- 512
DB 1963 -----TGKAFKGLNQIGDDKTYFNSDGVNMKGFGVSINDN 1996
```


[illegible]


```
Db      824 IYGD 827
RESULT 14
Q912M3 PRELIMINARY; PRT; 2468 AA.
AC Q912M3;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein PA1874.
GN PA1874.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales.
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warriner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Wu A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RA *Complete genome sequence of Pseudomonas aeruginosa PA01, an
RA opportunistic pathogen.*;
RT Nature 406:959-964(2000).
RL EMBL; AE004613; AAG05263.1; -.
DR InterPro; IPR001343; Hemlysn_Ca_bind.
DR InterPro; IPR006162; Psantne_attach.
DR PRINTS; PR00313; CABNDNGRPT.
DR PROSITE; PS00012; PHOSPHOPANTHETINE; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 2468 AA; 284414 MW; 13596AFAB2C4B899 CRC64;

Query Match 3.2%; Score 148.5; DB 16; Length 2468;
Best Local Similarity 19.5%; Pred. No. 2.3;
Matches 182; Conservative 90; Mismatches 301; Indels 361; Gaps 45;

QY 10 NTSQPINDLQSLAAEVKFAQSILP-----AHP-----KEGDSQPH----- 46
Db 11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 324 NSTSPVTEAPDTPAPADTVQVAPDGGSSVTGNAEPGATVGVDTDGQDPTTVVVGPG 383
QY 47 -----LTSLRKSLLLVRPVKADKTPQVVEARD-----DNNKIL----- 80
Db 11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 384 GSFEVPLNPLTNGETVTVITVDPAAGNSSTPTVTAEPDPPAPQVNASNGSVLSGTAEAG 443
QY 81 -----GTLTLYPPSSLPD-----TIYHLDPV 101
Db 444 VTIVITDGNPNIGOTSADANGNSFTPGSOLPDGTVVNVVARDAAAGNSSPATSTVDGV 503
QY 102 PEGGIDTFPHNGTKKIITVAEVNK---LSDASGSIHSHLTNVALVEIHTANGRWYDI 158
Db 11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 504 APNAPVVEPSNGSE---LSGTAEPGSSVTLTDGNGNPICQTAD-----ANGNW---S 550
QY 159 YLPOGPDLEGKMWRFVS---SAGYSS-----TFYGDGRKVTLSVGNLTLLFKYVNGWF 208
Db 551 FTFPTPLPDGIVNVVARDAAAGNSPPASVTVDAVAPATPTVDSNGTTL----- 600
QY 209 RSGELENRRITIAQHIWSAELPAHIVPLGLNVLIVKQGLNSGLRNDIKIAPGELLHTID 268
Db 601 -SGTAE-----PGSSVTLTDGN-----GNP----- 619
QY 269 IGMLTPTPRDFDAKDEKAHREYFOTIPVSRMIVNMYAPLHLKVMYLPTEGELLTD----- 323
Db 620 IGVTA-----DGSNGWTFETSTPLPNCVTNVNATATD-----PSGNASSPASVTV 664
QY 324 -----MDPGNGWISHTNRQIRIGELVSHGIDNANYGLNSTAGLGENSHPYVAQL 374
Db 665 DAVAPATPVVNPNSGTTLSGT-----AEPGATVTLTDG-NGNP---IQGV 705
```

```
QY 375 AAHNSRGNTA-----NGIQVH-----GGSGGGGIVTLDL-----TL 405
Db 706 TADGS-GNMSFTPTPLPNCVTNVNATATDASGNTSAGSSVTVDVAPATPVINPSNGTTL 764
QY 406 GNEFSHEVGHNYGLGHYVDGFKGSVHRSAAENNSNHWGDDKKRFIPNEVPS-QTNEKSC 464
Db 765 SG--TAEPGSSVTL---TDGNGNPICQVTDAGSGNW-----SFTPTPLADGTV 808
QY 465 LNNOCOEPPDGHKFGFDAMAGGSPFSAANRFMYTPNSSAIHQRFENKAVFDSRSSTGF 524
Db 809 VNATATDP-----AGNTSGOGSTVDGVAPTPTV-----NLNNGSSLSGTA- 850
QY 525 SKWNADTQEMEPYEHIDRAEQITASVNSLSESKMAELMAEYAVVYKVMHMGNGWTRNIYI 584
Db 851 -----EPGS-----TVILTDCNGNPVIAEVTAD-----GSGNWT---VT 880
QY 585 PTASADNRGSIILTINHEAGYNSYLFINGDEKVVYSGYKKSFYSDGQFWKRDVDTREAR 644
Db 881 PSTPIANGTVVNVVVAQDAAGNS-----SPGASVT-----VDSQAPA 916
QY 645 KPEQFGVPVTVTLVGYDPEGTLSYIYPAMYGYFTYSDSDSONLSDNDCQLQVDTKEG 704
Db 917 APVNVNSGTTLSGTAEFGATV-----TLTDGNGNPIC---QVTDG-GSGN 958
QY 705 LRFLANHRANNTVMNKFHINVPTESQPTOATLVCKNKKILDTKSLPAPEGLTYVNGQA 764
Db 959 WSFTPGTPLANGTVVN-----ATASDPTGNTSAPASTTVD--SVAPAAP-VVNPSNGAE 1009
QY 765 LPAKENECCIVSNNGKRYCLPYGQSGYSLPDW 798
Db 1010 ISGTAEPGATVTLTDGSGN--PIGVQVTDAGSGNW 1041

RESULT 15
Q8KRR3 PRELIMINARY; PRT; 3692 AA.
AC Q8KRR3;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Outer membrane protein.
OS Fusobacterium nucleatum.
OC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium.
OX NCBI_TaxID=851;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PK1594;
RA Ma X., Kempf M.J., Jewett A., Park H.-H., Shi W.;
RT *Cloning and analysis of Fusobacterium nucleatum apoptosis-inducing
RT genes.*;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF525505; AAM90995.1; -.
DR InterPro; IPR005546; Autotransporter.
DR Pfam; PF03797; Autotransporter; 1.
SQ SEQUENCE 3692 AA; 389826 MW; DFF61C00BA0ED180 CRC64;

Query Match 3.2%; Score 148.5; DB 2; Length 3692;
Best Local Similarity 18.8%; Pred. No. 4.3;
Matches 174; Conservative 106; Mismatches 319; Indels 327; Gaps 39;

QY 48 TSLRKSLLLRVPKADKTPQVVEARDNNKIL-----GTLTLYPPSSLPD 93
Db 162 SSVPTSLL-----PDKKTIINVSV--NNNRALVVDARENFDFQMKGTINLCNQNWGI 213
QY 94 TIY--HLDGVPEGID-----FTPHNGTKKIINTVAEVNKL-----SDASGS 133
Db 214 DLOGTHIGGSAKGSATNPAISTIRNEGVITGH-ATNKYDNNKATKEQVAFGFSNADASN 272
QY 134 SIHSHLTNNALVEIHTANGRWYDIYLPQGPDLBGMKAVRFVSSAGYSSTVFYGDQRK-VTL 192
Db 273 ATMTHTMINNEITLNPSSAGIQ--LKPEDP-----FYWDPNMGTL 311
```

Qy	193	SVGNTLLPKFYVNCQWFRSGELNNRITTYAQHIHSAELPAHWVPGI.NI.VIKOGNLSGRUN	255
Db	312	SSDNKMIINGLSG-----SRANNFGRVLMKADNR-----KDINL-----NGSGSGF	353
Qy	253	DIKIGAPCELLHTTIDIGMLTTPDRDFDFAKDEAHREYFQTIPTVSRMIVNNYVAPLHKE	312
Db	354	MIVTVNPGVIELNTISVNRITPTGSY-----NLRQNRIGT	390
Qy	313	VMLPTGEL-----LTDMDPGNGHWSHQMRQIRIGKELYSHGI--DNANYGLNSTAGLGENS	366
Db	391	KVLPGEICRSALSOSKTSYGVNSGNININ-GDESVGVGLHIEIOEVKIGGTINIGVES	449
Qy	367	HPYVVAQAAHNSRGNWANGIOVHGSGGGGIVTLD-----STLGNE-----FSH	411
Db	450	VSGTIGTSDTKSTTGDOTITLVNNAVGVFSGVPTPKVKNNEYDTMGNKNTGNIIGTETS	509
Qy	412	EVGHNYVGLG-HVYDGFKSGVHRSSEN-NNSWGWGDGDKKRFIPN---FYPQSOTNEKSL	465
Db	510	EYDGTINIGRHSKESIGLLVGDSETLNNGTLPNOARILKRSGSIYTKSSANNKLM	569
Qy	466	NNQCOEPE-----DGHK-----FGFDAMAGG	486
Db	570	NGNAVGVFKSDSNKSVFGSALDLOQSVDKTYWIGINKGNIDIGINSIGFVLLKGG	629
Qy	487	SPFSAA-----RFT-----MYTPNSAI	506
Db	630	NSKNTGNITVNESEDFWYPIVPTPNYDYRQNSIGFYGEODNETNEGTISVNTPNRSG--	687
Qy	507	QRFENKAVFDSRSSTGFSKKNADTQEME-----PYEHTIDRAEQITASVNE	553
Db	688	-----NKAVLLKGNNGITFNNTGDSVYKGRNNIAYAEKTYFNH-----EKNAAGTK	737
Qy	554	LSBSKAELMAEYAVYKVVWNGMNWTRIIYPTASDN-----RGS	594
Db	738	ISVGSNS-----IGIYVDKNTGVNIKAPIELADSNCTTIYGVSDGNAHINFENG	789
Qy	595	ILTINHEA-----GYSYLPIN--GDEKVVVSOGYKKSFVSDGQFVKWERDVTDR	642
Db	790	KLTIQKKAIGLFSQSTTNFKTEKFNENAGNELVNSDENSEAFA----PFGSGSDIAE	845
Qy	643	A-----RKPEOF---GVPTTLVGYDDPEGLTSSVIIPAMYGAGFYTSDSQN	688
Db	846	VLMKNIKFTAMKKGATFAVYVKGSIIVTLISHDFDTNTAKVYAPES-GTSVLVASDGS	904
Qy	689	LSDNDCOLQVDTKKEGOLR-----FRLAHRANNTVMNK	721
Db	905	OYDANKKLITNTNVGLVATNGSSNKSAINKGTIISKVNGVGLYANNNSEATNDIAG	964
Qy	722	PHINVPESOPTOATLVCHNKTILDTK	747
Db	965	-----ITMENKGSAAIIGENNSILTNK	986

Search completed: October 8, 2003, 16:27:23
Job time : 50.2838 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model
Run on: October 8, 2003, 16:20:57 ; Search time 21.7788 Seconds
(without alignments)
3912.317 Million cell updates/sec

Title: US-10-002-309B-2
Perfect score: 4709
Sequence: 1 MKLKYLSCTILAPLAIGVFS.....AWNGQYLDKPRSMRVYK 886

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 9616862 residues
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	4709	100.0	898	T42131	probable toxR-regu
2	2604.5	55.3	7	T00316	toxR-regulated lip
3	689.5	14.6	1002	T09438	toxR-activated lip
4	689.5	14.6	1013	B82276	ToxR-activated gen
5	671	14.2	1335	A82494	TagA-related prote
6	290	6.2	957	H82261	hemolysin-related
7	166.5	3.5	1217	F97177	alpha-glucosidase
8	161	3.4	1038	JC6027	115K outer membran
9	159.5	3.4	2367	S70172	toxin B - Clostrid
10	156.5	3.3	1157	F97255	fusion of alpha-gl
11	154	3.3	1211	S54500	alpha,alpha-trehal
12	153.5	3.3	2366	S10317	toxin B - Clostrid
13	148.5	3.2	2468	A83412	hypothetical prote
14	148	3.1	1021	I40805	collagenase - Clos
15	147.5	3.1	1530	AH1396	peptidoglycan anch
16	147	3.1	1377	D90538	hypothetical prote
17	147	3.1	1386	AC1533	surface protein (L
18	138.5	2.9	973	B86547	polymorphic outer
19	138.5	2.9	973	F72076	polymorphic outer
20	138.5	2.9	995	C81593	polymorphic membra
21	138.5	2.9	1268	B99789	hemagglutinin/hemo
22	138.5	2.9	1270	E85649	hypothetical prote
23	138	2.9	2894	C64474	hypothetical prote
24	137.5	2.9	2529	B64635	toxin-like outer m
25	136.5	2.9	3705	AD0123	probable autotrans
26	136	2.9	2167	AF1489	cell wall-associated
27	135	2.9	1579	B91290	probable invas
28	135	2.9	1700	G86131	probable invas
29	135	2.9	2051	S34688	enoyl-[acyl-carrie

ALIGNMENTS

RESULT 1
T42131

Probable toxR-regulated lipoprotein tagA - Escherichia coli plasmid p0157
C:Species: Escherichia coli
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Nov-2000
C:Accession: T42131; T00210
R:Burland, V.; Shao, Y.; Perna, N.T.; Plunkett, G.; Sofia, H.J.; Blattner, F.R.
Nucleic Acids Res. 26, 4196-4204, 1998
A:Title: The complete DNA sequence and analysis of the large virulence plasmid of Esc
A:Reference number: Z22068; MUID:98391744; PMID:9722640
A:Accession: T42131
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-898 <BUR>
A:Cross-references: EMBL:AF074613; PIDN:AAC70099; 1
A:Experimental source: strain EDL933; serotype O157:H7
R:Makino, K.; Ishii, K.; Yasunaga, T.; Hattori, M.; Yokoyama, K.; Yatsudo, H.C.; Kubo
S.; Shinagawa, H.
DNA Res. 5, 1-9, 1998
A:Title: Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an enterohemol
A:Reference number: Z14127; MUID:98290540; PMID:9628576
A:Accession: T00210
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 569-898 <NAK>
A:Cross-references: EMBL:AB011549; MID:d1204561; PIDN:BAA31757; 1; PID:d1032718
A:Experimental source: strain EHEC O157:H7, substrain RMD 0509952
C:Genetics:
A:Genome: plasmid p0157
A:Note: L7031

Query Match	100.0%	Score	4709;	DB	2;	Length	898;
Best Local Similarity	100.0%	Pred. No.	9.3e-297;				
Matches	886;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
QY	1	MKLKYLSTILAPLAIGVFSATAADNNSAIYFNTSQPINDLQSGSLAAEYKFAQSQILPAH	6C				
Db	13	MKLKYLSTILAPLAIGVFSATAADNNSAIYFNTSQPINDLQSGSLAAEYKFAQSQILPAH	72				
QY	61	PREGDSQPHLTSLRKSLLLVPRVKADDKTPVOVEARDNNKILGTLTLPSPSSLPDTIYH	120				
Db	73	PREGDSQPHLTSLRKSLLLVPRVKADDKTPVOVEARDNNKILGTLTLPSPSSLPDTIYH	132				
QY	121	LDGVEGGIDFTPHNGTKKIINTVAEVNKLSDAGSSISHSLTNALVEIHTANGRWVRD	180				
Db	133	LDGVEGGIDFTPHNGTKKIINTVAEVNKLSDAGSSISHSLTNALVEIHTANGRWVRD	192				
QY	181	IYLPQGPDLLEGKMWRFVSAGYSSVTFYGDRTVLSVGNLTLLFKYVNGQWFRSGELENR	240				
Db	193	IYLPQGPDLLEGKMWRFVSAGYSSVTFYGDRTVLSVGNLTLLFKYVNGQWFRSGELENR	252				
QY	241	ITYAQHNSAELPAHWIIVPGLNLVTKQGNLSGRINDIKTGPAGCELLHTIDIGMLTTPRD	300				

```
|||||
253 ITYAGHINSAELPAHWIIVPGLNVLKOGNLGRLNDIKIGAPGELLHTIDIGMLTTPRD 312
QY 301 RFDPAKDEAHREYFQIPVSRMIVNNAAPLHLKEVMLPTGELLTDMDPGNGHSGTMR 360
Db 313 RFDPAKDEAHREYFQIPVSRMIVNNAAPLHLKEVMLPTGELLTDMDPGNGHSGTMR 372
QY 361 QRIKELVSHGIDNANGLNSTAGLGNHSHYVVAQAAHNSRGNANGIQVHGSGGG 420
Db 373 QRIKELVSHGIDNANGLNSTAGLGNHSHYVVAQAAHNSRGNANGIQVHGSGGG 432
QY 421 IVTDLSTLGNESHEVGHNYGLGHYVDFGKGSVHRSANNSTWGWGDKKRFTIPNYP 480
Db 433 IVTDLSTLGNESHEVGHNYGLGHYVDFGKGSVHRSANNSTWGWGDKKRFTIPNYP 492
QY 481 OTNEKSLNNOCEPFDGKHFGFDAMAGGSPFSAANRFTMTTPNSSAIIOFFENKAV 540
Db 493 OTNEKSLNNOCEPFDGKHFGFDAMAGGSPFSAANRFTMTTPNSSAIIOFFENKAV 552
QY 541 SRSSTGFSKWNADTOEMEPYEHTIDRAEQITASVNELSESKAELMAEYAVVYVMNGN 600
Db 553 SRSSTGFSKWNADTOEMEPYEHTIDRAEQITASVNELSESKAELMAEYAVVYVMNGN 612
QY 601 WTRNIYIPTASADNRGSLTINHEAGYNSYLFINGDEKVVSGYKKSFPVSDGQFWKRDV 660
Db 613 WTRNIYIPTASADNRGSLTINHEAGYNSYLFINGDEKVVSGYKKSFPVSDGQFWKRDV 672
QY 661 VDTREARPEQGVPTLVGYDPEGLTSSYIYPAMYGAYGFTYSDSQNLSDNDQCLQ 720
Db 673 VDTREARPEQGVPTLVGYDPEGLTSSYIYPAMYGAYGFTYSDSQNLSDNDQCLQ 732
QY 721 VDTREGQLRFLANHRANTVNNKPHINVPTEOPTQATLVNCKNILDTSKLTAPAEGLT 780
Db 733 VDTREGQLRFLANHRANTVNNKPHINVPTEOPTQATLVNCKNILDTSKLTAPAEGLT 792
QY 781 YTVNGQALPAKENECEIIVSVNGRRYCLPVGQSGYSLPDWIVGQEVYDVSQAKAVLLS 840
Db 793 YTVNGQALPAKENECEIIVSVNGRRYCLPVGQSGYSLPDWIVGQEVYDVSQAKAVLLS 852
QY 841 DWDNLNRYNRIEFVGNVNPADMKKVKANNGOYLDPSKPRSMRVYK 886
Db 853 DWDNLNRYNRIEFVGNVNPADMKKVKANNGOYLDPSKPRSMRVYK 898

RESULT 2
T00316
toxR-regulated lipoprotein tagA - Escherichia coli plasmid p0157
C:Species: Escherichia coli
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 11-Jan-2002
C:Accession: T00316
R:Makino, K.; Ishii, K.; Yasunaga, T.; Hattori, M.; Yokoyama, K.; Yatsudo, H.C.; Kubota, S.; Shinagawa, H.
A:Title: Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an enterohemorrhagic E. coli O157:H7.
A:Reference number: 214127; MUID:98290540; PMID:9628576
A:Accession: T00316
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-587 <NA>
A:Cross-references: EMBL:AB011549; NID:d1204561; PIDN:BAA31835.1
A:Experimental source: strain EHEC O157:H7, substrain RMD 0509952
C:Genetics:
A:Gene: tagA
A:Genome: plasmid p0157

Query Match 55.3%; Score 2604.5; DB 2; Length 587;
Best Local Similarity 96.8%; Pred. No. 9.1e-161;
Matches 491; Conservative 2; Mismatches 7; Indels 7; Gaps 2;

QY 52 AQSQILPAHPKEGDSOPHLTSLRKSLLVLRPKADKTPVQVEARDNNKILGTLTLYPP 111
Db 63 AASQILPAHPKEGDSOPHLTSLRKSLLVLRPKADKTPVQVEARDNNKILGTLTLYPP 122
```

```
112 SSLPDTIYHLDGVEGGIDFTPHNGTKKIINTVAEYNKLSDSAGSSISHSLTNALVEIH 171
Db 123 SSLPDTIYHLDGVEGGIDFTPHNGTKKIINTVAEYNKLSDSAGSSISHSLTNALVEIH 182
QY 172 TANGRWVRDIYLPQGPDLGKMYRFVSSAGYSSTVFYGDVKVTLVSGNTLLFKYNGQWF 231
Db 183 TANGRWVRDIYLPQGPDLGKMYRFVSSAGYSSTVFYGDVKVTLVSGNTLLFKYNGQWF 242
QY 232 RSELENNRITYAQHTWSAELPAHWIIVPGLNVLKOGNLGRLNDIKIGAPGELLHTID 291
Db 243 RSELENNRITYAQHTWSAELPAHWIIVPGLNVLKOGNLGRLNDIKIGAPGELLHTID 302
QY 292 IGMLTTPRDRFPAKDEAHREYFQIPVSRMIVNNAAPLHLKEVMLPTGELLTDMDPGN 351
Db 303 IGMLTTPRDRFPAKDEAHREYFQIPVSRMIVNNAAPLHLKEVMLPTGELLTDMDPGN 362
QY 352 GGHSGTMRORIGKELVSHGIDNANTGLNSTAGLGNHSHYVVAQAAHNSRGNANGIQ 411
Db 363 GGHSGTMRORIGKELVSHGIDNANTGLNSTAGLGNHSHYVVAQAAHNSRGNANGIQ 422
QY 412 VHGSGGGGIVTLDSTLGNESHEVGHNYGLGHYVDFGKGSVHRSANNSTWGWGDKK 471
Db 423 VHGSGGGGIVTLDSTLGNESHEVGHNYGLGHYVDFGKGSVHRSANNSTWGWGDKK 482
QY 472 RFIPTNYPSTNKSCLNNOCEPFDGKHFGFDAMAGGSPFSAANRFTMTTPNSSAIIO 531
Db 483 RFIPTNYPSTNKSCLNNOCEPFDGKHFGFDAMAGGSPFSAANRFTMTTPNSSAIIO 542
QY 532 FEENKAVF----DSRSSTGFS---KVN 551
Db 543 FEENKAVSIIVPPASAGMQIRKKVN 569

RESULT 3
T09438
toxR-activated lipoprotein - Vibrio cholerae
N:Alternate names: toxR-activated gene A protein
C:Species: Vibrio cholerae
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 17-Nov-2000
C:Accession: T09438; A39108
R:Karaoz, D.K.R.; Johnson, J.A.; Bailey, C.C.; Boedeker, E.C.; Kaper, J.B.; Reeves, Proc. Natl. Acad. Sci. U.S.A. 95, 3134-3139, 1998
A:Title: A Vibrio cholerae pathogenicity island associated with epidemic and pandemic A:Reference number: 216672; MUID:98169509; PMID:9501228
A:Accession: T09438
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1002 <NA>
A:Cross-references: EMBL:AF034434; NID:g3004923; PIDN:AAC12274.1; PID:g3004926
A:Experimental source: strain N16961
R:Parsot, C.; Tauxem, E.; Mekalanos, J.J.
Proc. Natl. Acad. Sci. U.S.A. 88, 1641-1645, 1991
A:Title: ToxR regulates the production of lipoproteins and the expression of serum re A:Reference number: A39108; MUID:91156664; PMID:2000374
A:Accession: A39108
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-25 <PAR>
A:Cross-references: GB:M60658
C:Genetics:
A:Gene: tagA
A:Note: part of the pathogenicity island (VPI); associated with epidemic and pandemic C:Keywords: lipid binding; lipoprotein

Query Match 14.6%; Score 689.5; DB 2; Length 1002;
Best Local Similarity 24.2%; Pred. No. 2.1e-36;
Matches 232; Conservative 151; Mismatches 279; Indels 295; Gaps 39;

QY 27 NSAIYNTSOPINDLOGSLAAEYKFAQSILPAHPKEGDSOPHLTSLRKSLLVLRP-VKA 85
Db 137 NEIDYHTTEKKNYVGLGEYRFVQTHYI--SPEGRKNEPEIITGRDILIFKPSIKN 194
QY 86 DDKTPVOVEARDNNKILGTLTLYPPSSSLPDTIYHLDGVEGGIDFTPHNGTKKIINTVA 145
```

```

Db      195 SSSILMKIYSEDG---LTSKVMKSPMLPKTDQPID-----ID----- 230
Qy      146 EVNKLSDAGSSSTHSLTNNALVEIHTANGWRVRIYLPQDPDECKMVFYSSAGYSST 205
Db      231 ----- 230
Qy      206 VFYGDRAKVTLSVGNLTLLFKYVNGQWFRSGELENNR-ITYAQHIWSAELPAHRIWVPGNLV 264
Db      231 -----ENNVVSYSNSYSAEIPWKKMSGMSLH 259
Qy      265 IKOGNISGRU-----NDIKIGAPGELLHTIDIGMLTTPDRDFPAKDEKAHR-EYFQTI 318
Db      260 FEDEN--GNLGIIESERIKFSAPSELIIONIDGLMYPKPRGRNIVIKELERTAVDYFOKV 317
Qy      319 PVSRMIVNYPAPLHKEVMLPTCELLTMDPGNGGHSCTMRORIGKELYSHGIDNANYG 378
Db      318 PVSKLIFSDDTPHFKEKITLPNGTVITYTEKSAIDGGWHQDMREAVGKALVSTGINNANLG 377
Qy      379 LNSTAGLGENSHPYVVAQLAHNSRNYANGIOVHGSGGGGIVTLDSTLGNFSEHVG 438
Db      378 IVASSGYSO-QYNRLTNHITAHTNIGYNNGVVHGSGGGGIVTLENTLHNEMSHELGH 436
Qy      439 NYGLGHYVDFGKSVHRSANNNSTWGDGDKRRFIPNFPYPSOTNEKSCL--NNQCEPFP 496
Db      437 NYGLGHYVAG--GTSH---GPDTSWGDGYYKRFIANFDWKRSPQSNIRPDNQEVVYKPF 490
Qy      497 -DGHKFCFDMAGSSPF--SAANRFMTYTPNSNAILORFFENKAVFDSRSTGFSKNAD 553
Db      491 MDKITYLMDAMSGYDHQNGIISRYTLHHPYVARIIDMLKNGAVV---INNIDYMYND-E 546
Qy      554 TQEMEPYEHTIDRAEQITASVNELSEKMAELMAEYAVVHVHMGNGWTRNIYIPTASAD 613
Db      547 LKNIVYVYKGT-----NFKVPKKGVPVVTILGVYDPDKIN-----PSQLYPTYS-- 591
Qy      614 NRGSLITIN-----HEAGYN-----SYLFING-- 635
Db      592 NYGNIFOLEKPRSESSLGQYKVDVNYLDRVNTLHHTLVNKRKEKICFYSYSPRGKK 651
Qy      636 -----DEKVVYSGYKKSFSYDGFQWKEROVVDTRFARKPEQFVPTTLVG---Y 682
Db      652 FEFLGYEDIENKICTGGRSIHYLEDG---KKNPI-----ESKYNDYF---LLSIDGGEIS 701
Qy      683 YDPEGLT-----SSYIYPAAMYA-----YGFYSDSDSONLS-- 713
Db      702 YVPDSTIGESKICSLKMSGIVYGAGFIKNGSCQIDGVFMNGFOWAFTLNQSGVNSTYTW 761
Qy      714 DNDQLOVDYTKEGOLR-FRLANHRANVTYVANKPHINVPTEQPTQATLVNCKNILDTKSL 772
Db      762 SNECVLKIKDKDNKNNIESISIPNYRIEKNQSNKIHNLISREKPIIDINVCGEHELTSIKV 821
Qy      773 TPAPEGLTYTVNGQALPAKENEKGCIVSNKRYCLPVGQSGYSLPDWIVGQEVYVDSG 832
Db      822 SDNPD-----IKLLKGPPII-----VGQEHGYT-----SYEPKLPSPG 852
Qy      833 AKAKVLLSDMDNLS-YNRIGFEFVGNVAPADMKYKAWNGQYL---DFS-KPRSRVV 884
Db      853 -----WFKHYDNTPEKNEINHELKMRVND-----NDEYICRFNFSDDREMKFV 897

RESULT 4
ToxR-activated gene A protein VC0820 [imported] - Vibrio cholerae (strain N16961 serogic
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: B82276
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
charlson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, P.
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833; PMID:10952301
A:Accession: B82276

```

```

A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1013 <HE>
A:Cross-references: GB:AE004167; GB:AE003852; NID:9655268; PIDN:AAF93983.1; GSPDB-G
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC0820
A:Map position: 1

Query Match      14.6%; Score 689.5; DB 2; Length 1013;
Best Local Similarity 24.2%; Pred.No.2.le-36;
Matches 232; Conservative 151; Mismatches 279; Indels 295; Gaps 39;

```

```

Qy      27 NSAIYFNTSQPINDLOGSLAAEYKFAQSOILPAHPKRGDSQPHLTSLRKSLLLVRP-VKA 85
Db      148 NEIDIYHTTEIEKNYVLSGEVRFVQTHVI--SPGRKNEPEIITGRDALILFSPSINK 205
Qy      86 DDKTPVOVARDONNKILGTLLYLPSSLPDRIYHLDGVPEGIDPTPHNGTKKIINTVA 145
Db      206 SSSILMKIYSEDG---LTSKVMKSPMLPKTDQPID-----ID----- 241
Qy      146 EVNKLSDAGSSSTHSLTNNALVEIHTANGWRVRIYLPQDPDECKMVFYSSAGYSST 205
Db      242 ----- 241
Qy      206 VFYGDRAKVTLSVGNLTLLFKYVNGQWFRSGELENNR-ITYAQHIWSAELPAHRIWVPGNLV 264
Db      242 -----ENNVVSYSNSYSAEIPWKKMSGMSLH 270
Qy      265 IKOGNISGRU-----NDIKIGAPGELLHTIDIGMLTTPDRDFPAKDEKAHR-EYFQTI 318
Db      271 FEDEN--GNLGIIESERIKFSAPSELIIONIDGLMYPKPRGRNIVIKELERTAVDYFOKV 328
Qy      319 PVSRMIVNYPAPLHKEVMLPTCELLTMDPGNGGHSCTMRORIGKELYSHGIDNANYG 378
Db      329 PVSKLIFSDDTPHFKEKITLPNGTVITYTEKSAIDGGWHQDMREAVGKALVSTGINNANLG 388
Qy      379 LNSTAGLGENSHPYVVAQLAHNSRNYANGIOVHGSGGGGIVTLDSTLGNFSEHVG 438
Db      389 IVASSGYSO-QYNRLTNHITAHTNIGYNNGVVHGSGGGGIVTLENTLHNEMSHELGH 447
Qy      439 NYGLGHYVDFGKSVHRSANNNSTWGDGDKRRFIPNFPYPSOTNEKSCL--NNQCEPFP 496
Db      448 NYGLGHYVAG--GTSH---GPDTSWGDGYYKRFIANFDWKRSPQSNIRPDNQEVVYKPF 501
Qy      497 -DGHKFCFDMAGSSPF--SAANRFMTYTPNSNAILORFFENKAVFDSRSTGFSKNAD 553
Db      502 MDKITYLMDAMSGYDHQNGIISRYTLHHPYVARIIDMLKNGAVV---INNIDYMYND-E 557
Qy      554 TQEMEPYEHTIDRAEQITASVNELSEKMAELMAEYAVVHVHMGNGWTRNIYIPTASAD 613
Db      558 LKNIVYVYKGT-----NFKVPKKGVPVVTILGVYDPDKIN-----PSQLYPTYS-- 602
Qy      614 NRGSLITIN-----HEAGYN-----SYLFING-- 635
Db      603 NYGNIFOLEKPRSESSLGQYKVDVNYLDRVNTLHHTLVNKRKEKICFYSYSPRGKK 662
Qy      636 -----DEKVVYSGYKKSFSYDGFQWKEROVVDTRFARKPEQFVPTTLVG---Y 682
Db      663 FEFLGYEDIENKICTGGRSIHYLEDG---KKNPI-----ESKYNDYF---LLSIDGGEIS 712
Qy      683 YDPEGLT-----SSYIYPAAMYA-----YGFYSDSDSONLS-- 713
Db      713 YVPDSTIGESKICSLKMSGIVYGAGFIKNGSCQIDGVFMNGFOWAFTLNQSGVNSTYTW 772
Qy      714 DNDQLOVDYTKEGOLR-FRLANHRANVTYVANKPHINVPTEQPTQATLVNCKNILDTKSL 772
Db      773 SNECVLKIKDKDNKNNIESISIPNYRIEKNQSNKIHNLISREKPIIDINVCGEHELTSIKV 832
Qy      773 TPAPEGLTYTVNGQALPAKENEKGCIVSNKRYCLPVGQSGYSLPDWIVGQEVYVDSG 832
Db      833 SDNPD-----IKLLKGPPII-----VGQEHGYT-----SYEPKLPSPG 863

```

Qy 833 AKAKVLLSDWDNLNLS-YNRGCEFGVGNVPADMKKKYKAWNGQYL---DFS-KPRSMRVV 884
Db 864 -----WFKHYDNFEPKNEINHELGLKRVND-----NDEYICRFNFSOSDREMKFV 908

RESULT 5
A82494
TagA-related protein VCA0148 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: A82494
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qiu, H.; Dragoi, I.; Sellers, P.
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833; PMID:10952301
A:Accession: A82494
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1335 <HEI>
A:Cross-references: GB:AE004356; GB:AE003853; NID:99657536; PIDN:AAF96061.1; GSPDB:CN001
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VCA0148
A:Map position: 2

Query Match 14.2% Score 671; DB 2: Length 1335;
Best Local Similarity 25.5% Pred. No. 5.2e-35;
Matches 231; Conservative 108; Mismatches 248; Indels 318; Gaps 36;

Qy 39 NDQGSAAEVYFAQSOILPAHPK--EGDSQPHLTSL--RKSLLVRYKADKTPQVQE 94
Db 353 NLLQGSLEGLSITQHTSVA-PRGNELTQGHLDAMNREALLLTTPQOGEINQVRAE 411
Qy 95 ARDNNKILGTLIYPPSSLPDIYHLDGVPEGIDFTPHNGTKKIINTVAEVNKLSDAS 154
Db 412 VFLDGLVQTLML-PPSAL-----AASQDPENGRMKV----- 443
Qy 155 GSSTHSHTNNALVEIHTANGRWVRDIYLPQDPDLEKVRVYSSAGYSTVFYGRKVT 214
Db 444 ---VFSLA----- 449
Qy 215 LSVGNTLLFKYVNGQWFRSGELENNRITYAQHIWSAELPAHWIVPGLNLYI-----KQGN 269
Db 450 -----WSLPLOWDMKPGLSRLTLNLDNLDGREGV 476
Qy 270 LSGRLNDIKIGAPGELLHTIDIGLTTPRORDFAKD-KEAHREYFOTIPVSRMIVNXY 328
Db 477 LS---QGEIQFGGAPELVIQIDIGHMPPHDDRTNMTONLPTLAADYFQKIPASKLVNADY 534
Qy 329 APHLKEVMLPTGELLTMDPDGNGHSGTMRQRIKVELYSHGIDNANYGLNSTAGLGEN 388
Db 535 TPAHFVPTWPNGVVYTDKASSTGGHSGDMREAIKAKVSTGINNANVIVSSAGYSQQ 594
Qy 389 -----SHPYVVAQAAHNSRGNTAN-----GIQVHGGSGGGGIVTLDTLGNFESHEVG 437
Db 595 YNRRFNH-----ITAHTNVGIYTKKOTDLPOVVVHGGSGGGGIVTLEATTGNWSHEL 648
Qy 438 HNYGLGHVYDGFCKSVIRSNAENNNSTWGDGDKRPTPNF-----YPSQTNKESCLNNQ 491
Db 649 HNYGLGHW--PYMASIH-----DLESWGWDFAHFQRFIGNLFRWGDVYTOQQGD-----D 696
Qy 492 CQEPF-DGHKFGFDAMAGGSP--FSAANRTMTYTPNSSAIIORFEENKAVFDSRSSTGES 548
Db 697 IVPFKAERFLRQADGQGEQYVGTISRFTLEHPAQRKAQRMMNNGFNLDSPSGYV 756
Qy 549 KWNADTQEMPEYHTIDRAEQI-----TASYNELSES 580
Db 757 QWDQETORYKAVEYDTPKPOQGVVPVVTLLGIYDPQENPSQIPLYVSYNGVNFELPQP 816
Qy 581 KMAELMAEYAVYKVVHWMNGNNTNRIIYPTASADNRGSILTINHEAGVNSY--LFINDEK 638

817 EQGEYQLE-----GW-----QAQGD-----LT-QAEIQYNOWOTLLIDGQOL 852

639 VV-----SQYKKFSFVSDGQFWKERDYYD-TREARKPEQFGVPVTLVGYDPEGLTS 690

853 PICRFDYTWNGOSATFV--GSLNAQRNVCEGRDNR-----WYNDYQIDSPVG 899

691 SYIYPAMYGAYFTYSDDSQ-----NTSDND----- 716

900 QVELLSQFAGNVVTPNAEIGEIVQLCTLNKPHNNGSHDQAGFVRNGRCEQVEGVKNNAE 959

717 -----CQLOYDTKGQLRFLRNLHANHRANNTVMNKEHINVPT- 752

960 GRVWSYAINRSEVLSSTLASQRCELVVHRNGSHIALDGNRHKHSTESNKFHNLMSMEK 1019

753 SPTQATLVNCKNRIKLTSTPAEGLTIVNGQALPAKENEQECIVSVNSGRKYCLPVQ 812

1020 GVPQVSLSCSD-LNCTSLT-----RTPD-QNPPLDKLKGPII-----IQ 1060

813 RSGYS 817

1061 EYGS 1065

RESULT 6

H82261

hemolysin-related protein VC0930 [Imported] - Vibrio cholerae (strain N16961 serogrou

C:Species: Vibrio cholerae

C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001

C:Accession: H82261

R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.

chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers

l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A:Reference number: A82035; MUID:20406833; PMID:10952301

A:Accession: H82261

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-957 <HEI>

A:Cross-references: GB:AE004176; GB:AE003852; NID:g9655385; PIDN:AAF94092.1; GSPDB:GN

A:Experimental source: serogroup O1; strain N16961; biotype El Tor

C:Genetics:

A:Gene: VC0930

A:Map position: 1

Query Match 6.2%; Score 290; DB 2; Length 957;

Best Local Similarity 53.8%; Pred No. 1.5e-10;

Matches 57; Conservative 14; Mismatches 35; Indels 0; Gaps 0;

QY 779 LTVYNGQALPAKENEQECIVSVNSGRKYCLPVQGRSGYSLPDWIVGQYVDSGAKAVL 838

DB 13 LSLSSNVVQATTNEAEGCIISRLNGERYCLKVGRSGYSLPSNIYAHVPDVQAPSGVSM 72

QY 839 LSDWDNLNLYNRGEFVGNVPADMKVKAWNGQYLDPSKPSRNV 884

DB 73 LSDWDNLNLYNRGLAVFDRTYGNEDLNKYAYNGAYLDPSKPSRNV 118

RESULT 7

F97177

alpha-glucosidase fused to unknown alpha-amylase C-terminal. domain [Imported] - Clos

C:Species: Clostridium acetobutylicum

C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001

C:Accession: F97177

R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L

: Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium

A:Reference number: A96900; MUID:21359325; PMID:21359325

A:Accession: F97177

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1217 <KUR>

A:Cross-references: GB:AC001437; PIDN:AAK80209.1; PID:g15025253; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC2252

Query Match	3.5%	Score 166.5	DB 2	Length 1217
Best Local Similarity	18.6%	Pred. No. 0.021		
Matches	Conservative 121	Mismatches 319	Indels 343	Gaps 48
QY	79	LVRPVKA	---D	KTPVQVEARDNNKILGTLTLYPPSSLP-----DTIVHLDGVPPEGI 129
Db	29	IIRPVSAKAPNKKTIKTVENARV-----SNLSAKLNGDTLQVINGLDETDI 79		
QY	130	-----DFTPHNGTKKIINTYAEVNNKLSDASGSIHSHLTNNALV-----EIHHTA 173		
Db	80	KICEPOVLKVDYKP-SGSSSDTLVDPNKIWN-TGNIISDLNSDPWVITTKMTIKIS 137		
QY	174	NGRVRVDIYLPQGDLEGKMY---RFVSSAGYSTVFYDGRKVTLSVGNLTLLFFVYVQW 230		
Db	138	KSDLTMSVY-----DSTGKQIVQQSTASKSVFTHNSGDR-----FYGINGYN 181		
QY	231	FR-----SGELEN-NRITYAQHIWSAELPAJHIVPGLNLVKGQNLSGRLND----- 276		
Db	182	FKEDSNKGLMRNGTESYVACYQCHGSPFVMSNDGYGLLVDSOGSFTIGDTSLOYSGIS 241		
QY	277	-----IKTGARGELLHTIDI-----GMLTTP-----RORDFAKDKEAHRE 313		
Db	242	KTDTDYMLGSPKEVISESDVSGKAPMFPKWATGFTNTQGWNNNSISCTGNDEDLKS 301		
QY	314	YFOT-----IPVSRMIV-----NNYAPLHLKEYMLP---TGEILLTMDPGNGGWS 356		
Db	302	VLNTYRSKQIPIDNFCLDPEKKWGQDNYGEFFKNTDNFPAQNGOLKAYMD-SKGLKMT 360		
QY	357	GTRQRRI-----GKELVSHG-----IDNANTYGLNSTAGLGENSHPYV 393		
Db	361	GIMKPRILADSEQARYVTSKGWMLPGDASAADYCSGKMMENVFAISDV----- 409		
QY	394	VAQLAAHNSRGNTANGTOVHGSGGGGIIVLDTLGNFEFSHEVGHNYGLGHY----- 445		
Db	410	-----RKWMNNITQ-----DAPDKGIVGFWDCEDENVNFGFNMMNERA 450		
QY	446	VDGFKGSVHRSADNNNSTWG-----WDGDKKRFIPNFYPSQTNEX--- 485		
Db	451	IYDG-----QRAYKNQVRVSLNRYYAGQPYCYGMWSGD-----ISGFDPSMANQREM 500		
QY	486	SCLNQCQEPFDGKHKFGD--AMAGSP-----FSAANRFTMYTPNSSAIIQRFENKAV 538		
Db	501	LSAVN-----LGEAKGMGDTGGFNDGDPPTENYARWMEFSAFTP---IFRVHQDNRV 550		
QY	539	-----FDRSSTGFSKNADTQEMEPYEHTIDRAEQITA-----SVNELSEK 581		
Db	551	RYPWAFGSTAEAAKAKMOLRYTLIPYIYSDRSASGSLGLVRLSLMKEYPNDSNAANDX 610		
QY	582	MAELMAEYAVYKVHWNGNMTNRIYIP-----TASADNRGSILTIINHEAGYSY--- 630		
Db	611	EAMFEGDMLVSPVYVNGQTSKSIYLPBGNIDVTTGREYTGQ-OTINYAVDSTWNMSDIP 669		
QY	631	LFING-----DEKVSQGYKKSFVSD-----QQFWKER 658		
Db	670	LFIKSGAIIPTQDFENYVGEKKITDYYVADFPSPKATFTDYDDGTSYDENGYSYFPOK 729		
QY	659	DVVDTREARKQEQGVPTTLVLGYVDPGTLSSYI-----YP 695		
Db	730	MTLQTSIDTSKSVQENIDKNT--GSTYPD--LKDIYVKMYKGVNGAVTANGQALTOYSSVD 785		
QY	696	AMYGAYGFTYSDDSQNLSDNDCQLQVDTKEGQLRFRLANHRANNTVMKRFHINVTESQP 755		
Db	786	ALKSASGEGYASGT-----DTYGVWYVIKVSOGDAKN-----INVSCNPLP 826		
QY	756	TQATILVCNK-----ILDTKSLTPAPEGLT--YTVNGQALPAKENEGCIVSVNSKRYCL 808		
Db	827	VTITAAANPKGTYGYGPTQVTSKASKSDATIIYTL-----DGTPTVNSTK-VIA 875		

QY	809 PV 810
	:
Db	876 PI 877

RESULT 8
JC6027
115K outer membrane protein precursor - Bacteroides thetaiotaomicron
N/Alternate names: SusC protein
C/Species: Bacteroides thetaiotaomicron
C/Date: 31-Dec-1996 #sequence_revision 31-Dec-1996 #text_change 08-Oct-1999
C/Accession: JC6027
R/Reeves, A.R.; D'Elia, J.N.; Frias, J.; Salyers, A.A.
J. Bacteriol. 178, 823-830, 1996
A/Title: A Bacteroides thetaiotaomicron outer membrane protein that is essential for
A/Reference number: JC6027; MUID:96146534; PMID:850519
A/Accession: JC6027
A/Status: nucleic acid sequence not shown
A/Molecule type: DNA
A/Residues: 1-1038 <RES>
A/Cross-References: GB:L49338; NID:g1100064; PIDN:AAA95938.1; PID:g1100065
A/Note: It is uncertain whether Met-1, Met-14 or Met-20 is the initiator
C/Comment: This protein is an integral outer membrane protein, and it is essential fo
C/Genetics:
A/Gene: susC
C/Keywords: transmembrane protein
F:1-39/Domain: signal sequence #status predicted <SIG>
F:40-1038/Product: 115K outer membrane protein #status predicted <NAT>
F:1028-1038/Domain: transmembrane #status predicted <TM>

Query Match	3.48;	Score 161;	DB 2;	Length 1038;
Best Local Similarity	18.14;	Pred. No. 0.037;		
Matches 180;	Conservative 120;	Mismatches 328;	Indels 366;	Gaps 46;
Qy	39	NDLOGSLAAEVKFAQSOILPAHPEKGDSPHLTSLRKSLLVR---	PVRADDKTP---V	91
Db	131	NDMTGSVAA-----IKPDELSK-----ITNAQDMLSGKTAGVSVISNDGTPGGGA	177	
Qy	92	QVPEARDDNNKILGTLTYLPPSSLPDITYHLDGVEPGIDFTPHNGTKKIIINTVAEVNK--	149	
Db	178	QIRIRGGSSL-----NASNDPLIVIDGL--AID--NEGIKGMANGLSWMVPAD	221	
Qy	150	-----LSDASGSSIHSHLTNNALVEIHTANGRWVRDIYLPQGPDLGKMKVRFVSSAGYS	203	
Db	222	IETLVTKDASATAIYGSRASNGVIIITTKKG-----NGQAPSVTYNGSVFS	270	
Qy	204	STVFGDRKVTLSVGNLTLLFKYVNGOWFRSGELENNRITYAQHIWSAELPA-----HW	256	
Db	271	KT-----QKRYDVLSGDEVRA-----YANLWGDKLPAIDLGTANTDW	307	
Qy	257	-----IVPGLMLVIKQG--NLSGRBLNDIKIGAPGELLHTTIDICMLTTPPRDRFFPAK	306	
Db	308	QDQIFRTAVSTDHRVSYNGGFKNLPYRVS-----LGYTDODNGIVKTSNFR--RFTA	356	
Qy	307	DKEAHREYFO--TIPVSRMIV--NNKAPLHLKE---VMLPTGELLTMDMP--GNGG	353	
Db	357	SVNLAPSEFFEDHLKFNNAPFNCKNRYADSRCRYWRALADTPRPVYSNEDPYQFTGGY	415	
Qy	354	WHSCTMPORTGKELVSHGIDNANGLNSTAGLGE-----NSHPYVVAQLAHAHSRG	404	
Db	417	WQN-----INSTGFSNPDWKYTSNPNPSQPONPLAALELKNDKG	454	
Qy	405	NYANGIQVHGSGGGIVTLDSLTIGNFESHVGHNYGL-----GHVYDGFKG	451	
Db	455	N-----SNDVFGVNDVDFYKHFHFLPDLRLRLNASTIGETAEGTQT	491	
Qy	452	SVHRSAENNNSTWGDG---KKRFIPNEYPSOTNEKSLNNOCQEPFDGKHGFGDAMA	507	
Db	492	TIVSPYFGNNYWGNGDVQYKYNLSNIYVQYI--KSLGAND-----FDLMV	538	
Qy	508	GGSPFSAANRPTHYTPNSSAIIQRFFENKAVFDRSSTGFSK-----WNAADTQEMEPYEHTI	564	
Db	539	GGE-----QHFRN-----GFEGOGWDSYQ-----PHDA	565	

QY	565	DRAEQITASVNSESEKMA-----ELMAEYAVVVKVVMNWNTRNIYIPTASADNRGSILT	620
DB	567	KLREQTAYATRNLTLSYFGRLNYSLLNRYLFTFTKRWGGS-----SRFSKDNRW----	615
QY	621	INHEACYSNYLFINGDEKVVVSGYKKSFYSDGQFVKERDVVDTREARKEQFGVPVTILV	680
DB	616	-----GTFPSLAIGWK---IKREENFLKDVNVLSDLKLR-----LGMGTGQO	654
QY	681	GYVDPGEGTLSSYIYPAMYGAYGF--TYSDDS--QNLSNDMDCOLQVDTKEGQLRFRLANHR	736
DB	655	NIGDDFAYLPLVYNNEYAQYFPGDYIYSRSRPAFNEMLKWEKTTIWNAGLDGFLNGR	714
QY	737	ANNVANKHINVPTESQPTQATLVCNKILDTKSLTPAPEGLTY-----TVNGQ	786
DB	715	ITGGIDGYF-----RKMTCTVTAALRSPMNLQCPDDTEYRFTGKLRGYGSINAK	763
QY	787	ALPAKE-----NEGCIVSVNSG--KRYCLPVG-----	811
DB	764	PVTKDTWDLSYNTWNNHNETKLGGGDDSYVVEAGDKISRGNNTKVOAHKVGVAANS	823
QY	812	-----QRSGYSLDPIWIVGQYVVDGSAKARVLLSDMDNLSNRIGEFYGVN	858
DB	824	FVYSRGNTKVOAHKVGVAANSFYVYQVYDENGKPI-----ENMFYDRNGN--GTID	874
QY	859	PADMKVKVANG-----QYLFDSKPRSMR	882
DB	875	SGDKYIYKYPAGDVLMLGLTSKMQYKNFDFSFSLR	908

RESULT 9

S70172

toxin B - Clostridium difficile

C:Species: Clostridium difficile

C:Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 15-Oct-1999

C:Accession: S70172; S44271

R: von Eichel-Streiber, C.; Meyer zu Heringdorf, D.; Habermann, E.; Sartingen, S. Mol. Microbiol. 17, 313-321, 1995

A:Title: Cloning in on the toxic domain through analysis of a variant Clostridium difficile

A:Reference number: S70172; MUID:96079281; PMID:7494480

A:Accession: S70172

A:Status: preliminary; nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 1-2367 <VON>

A:Cross-references: EMBL:232277; NID:g761713; PIDN:CAA80815.1; PID:g761714

A:Experimental source: isolate 1470

R: Sartingen, S.; von Eichel-Streiber, C. submitted to the EMBL Data Library, July 1993

A:Description: Cloning and sequencing of an Clostridium difficile toxin B mutant.

A:Reference number: S44271

A:Accession: S44271

A:Molecule type: DNA

A:Residues: 1-1323, 'N', 1325-2367 <SAR>

A:Cross-references: EMBL:Z23277

C:Superfamily: cpl repeat homology

C:Keywords: cytotoxin

Query Match		3.4%;	Score 159.5;	DB 2;	Length 2367;
Best Local Similarity		18.4%;	Pred. No. 0.17;		
Matches	185;	Conservative	149;	Mismatches	321;
				Indels	353;
				Gaps	49;

QY 27 NSAIYFNTSQPINDOGLSAAEYKFAQSQILPAHPKEGDSQPHLTSLRKSLLLVRPVKAD 86 || DB 1454 NSELQKNIPYSFVDSSEK-----ENGFINSTKEG---- | 1501 |
QY 87 DKPT-----VQVEARDNNKILG-----	124
DB 1502 DSKPSPCYGSNNLKDVKVITKDNVNILTYGVLKDDIKISLSLTLQDEKTIKLSVHLD--	1559
QY 125 PEGGIDTPHNGTKKIINTVAEYVKNKLSADSGSSISHSLTNALVEIHTANGRWVROIYLP	184
DB 1560 -ESG-----VAEILKFNRRKS---TNTSDLSMFSLESM---IKSIFV-	1596

QY	185	QCPDLGKAVRVSSAGY--STVYFG-----	DRKVTLSVQNT	220
		: : : :		
DB	1597	---NFLQSNIFILFDANFIISGTSIGGFEFICDENNNIQPFYKFNFTLETNYTLVGNR	1653	
QY	221	L-LFKYVNGOWFSGELENNRIYQAQ-HIWSAELPAHWIPLVGLNLVIAKQGNLSG----	272	
DB	1654	QNKVIEPNYDLDDSGDISTVINFQKLYIGYDSCVKNKVVISPNYITDEINITPYETNN	1713	
		: : : : : : : : :		
QY	273	-----RLNDIKI-----GAPCELLHTIDIGMLTTPDRDF-DFAK	306	
		: : : : : : : :		
DB	1714	TYPEVIVLDANYINEKINYNIINDLSTRYVMSDGNDFILMSTSEENKYSQVKIRVNVFK	1773	
QY	307	DKE-AHREYF-----QTIPVSRMIVNNAPLHLKEVML-----	338	
		: : : : : : : :		
DB	1774	DKTLANKLSFNFSRQDQVPVSEIIL-SFTPSYIEGLGIDGLVSLVNEKFIYINFCMM	1832	
QY	339	-----PTGELLTD-----MDPGNGWHSCTMRIGICKELVSHGI	372	
		: : : : : : : :		
DB	1833	VSGLIIVINDSLVYFKPPVNNLTITFTVVGDDKYFENPINGGAAS-----IGETI----	1882	
QY	373	DNANYCLNSTAGLGENSHPYVVAQLAAHSRGNY-----ANGIQVHGSGGGGIWLDST	427	
		: : : : : : : :		
DB	1883	DKKNYFNGSG-----VLQTCVFSTEDGFIYFAPAN-----TLDEN	1918	
QY	428	LGNE-----FSHEVGHNTGLGYVDGFKGSHVRSAAENNNSTGW-----DGOKKRFIPNFYSQ	481	
		: : : : : : : :		
DB	1919	LEGEAIDFTGKLIIDENIYYFEDNYRGAYE-----WKELDGEHMYFSPE-----	1962	
QY	482	TNEKSLNNOCEPDPDGHKFGFDAMAGGSFSAANRF--TMYTPNSSAIQRF-----EN	535	
DB	1963	-----TGKAFKGLNQIGDDKYFNFSOGVMOKGFVSINDN	1996	
QY	536	KAVFDSKS--STGFSK-----WNADQEMEPVEHTIDRAEQITASVNELSKKAELMA	587	
		: : : : : : : :		
DB	1997	KHYFDDSGYKVGYYTIDGKHFFYAENGEMQGVFNTDGFYFAHHNEDLNGESEBIS	2056	
QY	588	EYAVVKVHMNGWNTRIYIPTASADNRGSIITINHEAGYNSYLFYNGDEKVVSGYKKS	647	
		: : : : : : : :		
DB	2057	YSGIL-----NFNKKIYV--FDDSFYAVGVGKLEDCSKYYF--DEDTAEAYIGLS	2103	
QY	648	FVSDGFQWKERDYYVDTRAKRPFQFQ--VPVTTILVGYDPEGTLSYYIYPAMYGAYFTYS	706	
		: : : : : : : :		
DB	2104	LINDGQYFNDGI-----MQGVFTINDKVFYFSDSGIIES-----GVQNI	2145	
QY	707	DDSONLSDNDQIQV---DTKEGQLFRLANRANNTVMKPHINVPTESOPTQATLVCN	763	
		: : : : : : : :		
DB	2146	DDNYEYIDONGIVQIGVEDTSDGKYFAPA-----NTVNDNIY-----	2183	
QY	764	NKILDTKSLTPAPE-----GLTIYVNGQALPAKENEGCIVSYNSGKKRYCLPVGQRS--GY	816	
		: : : : : : : :		
DB	2184	GQAVEYSGLVRVGEDVYVYGETTITGWIYDKENE-----SDKIYFVPETKKACKGI	2236	
QY	817	SLPDWIVGQVYVYDYGAKARVLLSDWINDLSY--NRIGEF-VGNVNPAD	861	
		: : : : : : : :		
DB	2237	NLIDDI--KYFDEKGIKMTGLISFENNYYENGEIQFGVINIED	2281	

RESULT 10
 F97255
 fusion of alpha-glucosidase (family 31 glycosyl hydrolase) and glycosidase (Trea/Mals
 C:Species: Clostridium acetobutylicum
 C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
 C:Accession: F97255
 R:Noiling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L
 :; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
 J. Bacteriol. 183, 4823-4838, 2001

RESULT 10

F97255

fusion of

C: Species

C; Date: 1/

C: Accession

R; Nolling.

∴ Dady, E

J. Bacter.

A;Title: C

A; Referenc

A; Accessio

A; Status:

A; MOLECULES

A; RESIDUES
A: CROCODILES

A; Experimental source: Clostridium acetobutylicum ATCC824

C; Genetics:

A; Gene: CAC2891

```
Query Match      3.3%; Score 156.5; DB 2; Length 1157;
Best Local Similarity 19.3%; Pred. No. 0.086;
Matches 198; Conservative 120; Mismatches 335; Indels 371; Gaps 56;

QY 83 VKADDTPTVOVEARDNNK----ILGTLTLYPPSSLPDTIYHLDDGVPEGGIDFTPHNGTK 138
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 35 VKADTKVKNKNSKSSQKFKRAKLKNTLTKYKGDDEIIRICEPQVKVDYKP-NGKS 93

QY 139 KIINTVAENKLSDAGSSSHSLTNALVEIHTANGRWVRI-----YLPQGPDL 189
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 94 SKOTLVVDPNK-KWSTGNIYSSDIKSDPMV---ITTKMYLKKINKEDLSILVY-----DL 144

QY 190 EGRWRFVSSAGYSSIVF---YGRKVTLSVGNWTLTLPKYVNGQWFR---SGLEN-NRI 241
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 145 QGKLLKQDSTASTASFTHNSGDR-----FYGINGYNFQEDSKGLMRNGTES 193

QY 242 TYAQHTWSAELPAHWIWPGLNLVIKQNTLSGRLND-----IKIGAPGEL 285
   || : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 194 VIAGYQCHGCSFPWSDGGLLVDSOGSFTIGDTSKLYDGISKDTDYVVMVGNPKKEI 253

QY 286 LLHTIDIG-----MLTTPDRF-----DFAKDEAHREY-FQIPVSRMI 324
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 254 LSESDVSGRAPFPKFWANGFTNTQMGWNSLSGTGNDEAKLSVINTYRSKOLPIDNFC 313

QY 325 V-----NNYAPLHLKEVMLP---TGELLTMDPGNGGWHSGTMRQRI-----GKE 366
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 314 LDFDMKKWGQDNYGEFKWNTDNFDSONGOLKAYMD-SKGLKMTGIMKPRILADSKOGRY 372

QY 367 LVSHG-----IDNANYGLNSTAGLGENSRHPYVVAOLAHAHNSRGNYANG 409
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 373 VTSKGWLPDGSSEADYCSCKMENVNFAL-----PQV-----RKWNNN 412

QY 410 IQVHGGGGGGIVTLDSITLANEFSHEVGHNYGLGHY-----VDGFGSKVHRSANEN 460
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 413 IQ---GAFDKGIVGF-----WNDECDENVNFGNMMNERAIYDG-----QRRHKN 456

QY 461 NSTWG-----WQDCKRFIPNYPQSOTNEK-----SCLNNQCQEPEDGHK 500
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 457 QRWVSLNRNTYAGAQRYSGYGMWSD-----ISTGFDMSANQRERMLSAVN-----LGEAK 506

QY 501 FGFD--AMAGSP-----FSAANRFTYTPNSSALIORFF--ENKA-----VFDGRSSTGFS 548
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 507 WGHDTGFGNGDPTPENYARWMEFSAFTP-----IFRVHGDKNKYRYPWAFGSTAEATAK 561

QY 549 KWNADTOEMEPYEHTIDRAEOITA-----SVNELSESKMAELMAEYAVKVVHM 596
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 562 KAMQLRYTLIPYIYSYDRSASQSLGLVRSLLMMEYPNDSNAANDKEAWFGDYMLVSPVY 621

QY 597 WNGWNTNRIYIP-----TASADNRGSILTIINHEAGYNSY---LFLNG----- 635
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 622 QEGTSSKSYLPEGNWIDYTTGREITGG-QTINTAVDSTNWSIDPLFIKSGAIPTODFE 680

QY 636 ---DEKVVYSGYKASF-----VSDGQFWKRDVVDYREARKPQFG 673
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 681 NYVGEKKITDYVDVDAFPGNEASSFDYDDGTSYNYENGSYFDQKMTLERAKDLKSVQFN 740

QY 674 VPVTTLVGYDPEGTLSYIYP-----AMYGAGFTYSDDSQNLSNDNC 717
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 741 ISPXT--GY--KSDLANYIVKHHKVSQGDVTGVRRIYASY-----DELKNAQGEY 791

QY 718 QLOYDTREGQLRFL--ANHRANTVMKFFHINVPTESQPTQATLVCKNKL-----DTKSL 772
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 792 VVGDDTYGVSYYIIVKVSAGHDKN-----INVP-----CNOVQLTAYADYK-- 830

QY 773 TPAPEGLTYVNGALPAKEN-----EGCIVSVNSCKRYCLPVGORSQSLPDI 822
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 831 -----GGTYT--SPQKSVLSKADPNAAIYYTLDTGTAFTVNSTK-----YTGPF--- 870

QY 823 VGEQVYVDSGAKAKVLLSDMDNLSYNRIGEPVGNVNPADMKKVKVXAWNGQYILDFSKPRSMR 882
```

```
Db 871 ----ITIDSSSKTLKFIVRD-----ANGNESDV-----FTEQYTTY-----IK 903
QY 883 VVYK 886
   |||
Db 904 VHYK 907

RESULT 11
S54500
alpha.alpha-trehalase (EC 3.2.1.28) - yeast (Saccharomyces cerevisiae)
N;Alternate names: acid trehalase ATH1; protein IP9367.06; protein IPR026w
C;Species: Saccharomyces cerevisiae
C;Date: 08-Jul-1995 #sequence_revision 19-Oct-1995 #text_change 29-Oct-1999
C;Accession: S54500; S60459; S70514
R;Badcock, K.; Churcher, C.M.
submitted to the EMBL Data Library, May 1995
A;Reference number: S54059
A;Accession: S54500
A;Molecule type: DNA
A;Residues: 1-1211 <BAD>
A;Cross-references: EMBL:249274; NID:g809585; PIDN:CAA89280.1; PID:g809591; MIPS:YPR
A;Experimental source: strain AB972
R;Desruelle, M.; Holzer, H.; Klionsky, D.J.
Yeast 11, 1015-1025, 1995
A;Title: Isolation and characterization of a novel yeast gene, ATH1, that is require
A;Reference number: S60459; MUID:96076626; PMID:7502577
A;Accession: S60459
A;Molecule type: DNA
A;Residues: 131-1167, 'FAG', 1171 <DES>
A;Cross-references: EMBL:X84156; NID:g1061283; PIDN:CAA89961.1; PID:g1061284
A;Note: the authors translated the codon GTT for residue 18 as Leu, GAG for residue
R;Alizadeh, P.; Klionsky, D.J.
FEBS Lett. 391, 273-278, 1996
A;Title: Purification and biochemical characterization of the ATH1 gene product, vac
A;Reference number: S70514; MUID:96350535; PMID:8764988
A;Accession: S70514
A;Molecule type: protein
A;Residues: 368-369, 'X', 371-377, 'X', 379, 'X', 381, 385-392, 'X', 394, 'X', 396 <ALI>
C;Genetics:
A;Gene: SGD:ATH1
A;Cross-references: SGD:S0006230; MIPS:YPR026w
A;Map position: 16R
C;Function:
A;Pathway: trehalose degradation
C;Keywords: glycoprotein; glycosidase; hydrolase; transmembrane protein; yeast vacuo
F;51-67/Domain: transmembrane #status predicted <TM>

Query Match      3.3%; Score 154; DB 2; Length 1211;
Best Local Similarity 18.4%; Pred. No. 0.13;
Matches 143; Conservative 104; Mismatches 265; Indels 266; Gaps 35;

QY 97 DDNKKILGTLTLYPPSSLPDTIYHLDDGVPEGGIDFTPHNGTKKIINTVAENKLSASGS 156
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 127 DDENMLIGS-NLFSKNTYSRQPYVANGYIGSR!---PNIGFGYALDT---LNEYTDAPGA 179

QY 157 SIHSLTNNALVEIHTANGRWVRIYLPQGPDLGKMRVFSVSSAGYSSTVFYGDRTVLS 216
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 180 -----LNNGWPLNRHFRFAGAFVSDFTCLQ-PKLSNFINPELDDVGYSTVI-----SSIP 227

QY 217 VGNLTLLFKYVN-QGTFRSGELENNRIT-YAQHTWSAELPAHWIWPGLNLVIKQNTLSGRL 274
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 228 QWNTLQPSLVNDSKWFNPQVTLDDVTNYSQ-----NLSMKDGIIVTTEL 271

QY 275 N--DIKIGAPGELLH-----TIDIGMLT--TPRD-----RPFDAKDEAHREY 314
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 272 DMLNSQIHVAKSEIWAHRHIIPLGVWSLETSNLDHLPDSFDSLDVNIWDLIDFNTSHTV 331

QY 315 FQT-----IPSRMIVNNAAPLHLKEVMLP--TGELLTMDPGNGGW 354
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 332 LHSTGTDEKKNVFMIVOPDNVFPSSKCAIYSTCTVYENSTNFINSESFEEKDVSSNIY 391

QY 355 -----HSGTMRQRIGKELVSHGIDNANYGLNSTAGLGENSHPTVVVAQLAA 399
```



```
Db 392 NVLTEDQPKIIVHKYVGMSTFKNKEQ--DNTNIGL      |||||      :|||
Qy 400 HNSRCNYANGIQVH-----                    :|||
Db 435 LNSKGNTEKLLSSHKRAWYDIYNDAFIEIPSDSLEMTAKSSLHLLANTRDYNVSSDRG 494
Qy 414 ---GGSG---GGGIVTLDSTLGN-----FSHEVGHNTGLGHYVDGFGKSVHRSANN 460
Db 495 LPVGVSGLSDSDYSGMFWFDADIWKEPALLPFPNVAQN-----MNNYRNATHSQAKLN 548
Qy 461 NSTWGWGDKKRFTPNFYPSOTNEKSLNNOCCQPFPGHKGFGFDAMAGGS-----510
Db 549 AEKYGYPGAIYPTSGKYANTCTGPGVDYEHYHNVDMAMASFYILNGHEGIDDEYLRY 608
Qy 511 --PF--SAANRFTMYTP--NSSAIQRFEE-----NKAVFDSRSSTGFSKNAD 553
Db 609 TFWPIIKNAOFFTAYVAYKNSLSGLYETYNLTDPDEFANHINNAFTNAGIKTLKWKATD 668
Qy 554 TQEMEPYEHTIDRAEQITASVNELSESKMAELMAEYAVVYHVMWNGWNRNIYIPTASAD 613
Db 669 -----IGNHLGEVDPKWSEI-----SKDIYIPRSSN 696
Qy 614 NRGSLITINHEAGYNSYL-----FINGDEKVVSGYKFSVSDQFQWKR 658
Db 697 -----ITLEY--SGMNSSVEIKOADVTILMVPYPLGYIN--DESILNNAIK-----DLYYYSER 744
Qy 659 DVVDTRARKPEQ--FGVPTTLVGYDPEGTLSVYIPAMYGAFTYVSDSDSNLSDN 715
Db 745 -----QSASGPMYTFVFAAAGLLNHGSSQSLYKSVLPYLRAPFAQFSEQSDDN 797

RESULT 12
S10317
toxin B - Clostridium difficile
C:Species: Clostridium difficile
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 15-Oct-1999
C:Accession: S10317; S21894; S22434
R:Barroso, L.A.; Wang, S.Z.; Phelps, C.J.; Johnson, J.L.; Wilkins, T.D.
Nucleic Acids Res. 18, 4004, 1990
A:Title: Nucleotide sequence of Clostridium difficile toxin B gene.
A:Reference number: S10317; MUID:90326540; PMID:2374729
A:Accession: S10317
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-2366 <BAR>
A:Cross-references: EMBL:X53138; NID:g40442; PIDN:CAA37298.1; PID:g40443
R:Eichel-Streiber, C.; Laufenberg-Feldmann, R.; Sartingen, S.; Schulze, J.; Sauerborn, M.
submitted to the EMBL Data Library, July 1991
A:Description: Comparative analysis of Clostridium difficile toxins A and B.
A:Reference number: S21894
A:Accession: S21894
A:Molecule type: DNA
A:Residues: 1271-2366 <EIC>
A:Cross-references: EMBL:X60984; NID:g40445; PIDN:CAA43299.1; PID:g40446
R:von Eichel-Streiber, C.; Laufenberg-Feldmann, R.; Sartingen, S.; Schulze, J.; Sauerborn, M.
Mol. Gen. Genet. 233, 260-268, 1992
A:Title: Comparative sequence analysis of the Clostridium difficile toxins A and B.
A:Reference number: S22434; MUID:92293124; PMID:1603068
A:Accession: S22434
A:Molecule type: DNA
A:Residues: 1791-2366 <VON>
A:Cross-references: EMBL:X60984
C:Genetics:
A:Gene: toxB
A:Superfamily: cpl repeat homology
C:Keywords: cytotoxin
```

Query Match 3.3%; Score 153.5; DB 2; Length 2366;

Best Local Similarity 18.6%; Pred. No. 0.41;

Matches 186; Conservative 155; Mismatches 321; Indels 339; Gaps 51;

Qy 27 NSAIYFNTSQPINDLQGLAAEVKFAQSIQLPAHPKEGDSQPHLTSLRKSLLLVRPVKAD 86

```
Db 1453 NSELQKNIPYSFVDSSEK-----ENGFSINGSTKEG-----LFVSELDPDWLISKVYMD 1500
Qy 87 DKTP-----VQVEARDNNKILG-----TLTLYPPSSLPDIYHLD-- 122
Db 1501 DSRPSFGYSSNNLKDKVITKDNVNILTYGLYKDDIKISLTLQDEKTIKLSVHLD 1560
Qy 123 GVPESGGIDFTPHNGTKIINTVAEVNKLSDASGSSIHSLTNALVEIHTFANGRWYRD-- 180
Db 1561 GVAE-----ILKFMKRGNTN--TSDSLASFLESNNIKSIFVNFLOSNIKFIILDAN 1609
Qy 181 -----IYLPQDPDLEGMVRFVSSAGYSTVEYDKRVTL-----S 216
Db 1610 FIISGTSISQFECIDENDNIQPIFKF--NTLETNYTLVGNRQNNMIVEPNYDLDSDG 1668
Qy 217 VGNLTL---PKYVNGQWFRSGELEN--NRITYAQHWSAEL---PAH-----WLPVG 266
Db 1669 ISSTVINFSQILY-----IDSCVNKYVISPNYITDEINITPVYETNNYTPYEVILVD 1721
Qy 261 LNLVIKQGNLSGRINDIKI-----GAPGELLHTIDIGMLTTPRORF--DFAKDKE--AHR 312
Db 1722 ANYNEKINYN--INDLSIRVMSNDCNDFILASTSEENKVSQVKIRFVNVFKDKTLANK 1779
Qy 313 EYF-----QPIPSRMVNNYAPLHLKEYML-----336
Db 1780 LSFNFSQKQDVPVSEIIL--SFTPSYEDGLGYDLGLVSLYNEKFYINNFMMVSGLIYI 1838
Qy 339 -----PTGELLTD-----MDPNCNGWHSQTHMRQIRIGKELVSHGIDNANYGL 379
Db 1839 NDSLYYKPPVNNLITGFVTVGDDKYFFNPINGAAS-----IGETI-----IDDRNYF 1888
Qy 380 NSTAGLGENSEBPVYVAQAAAHNSRGNY-----ANGIOVHGSGGGGIVTLDSTAGNE-- 431
Db 1889 NQSG-----VLQTVGFSTEDGPKYPAPAN-----TLDENLEGAID 1924
Qy 432 FSHEVGHNYGLHYVDGFGKSVHRSANNSTWGW---DCKKRFIPNFPSTQNEKSCL 488
Db 1925 FTGLKIITDENIYYFDNRYGAVE-----WKELDGEHMYFSPE-----1961
Qy 489 NNQCOEPFDGKHGFDAMAGGSPESAARF--TWYTPNSSAITORFF---ENKAVFDSR 542
Db 1962 -----TGKAFKGLNQIDGYKYFYNSDGVYKQKGFVSNIDNKNHFDD 2002
Qy 543 S--STGFSK-----WNADTQEMEPYEHTIDRAEQITASVNELSESKMAELMAEYAVKY 594
Db 2003 GVMKVGYTEIDGKHFFYFAENGEMQIGVFNTEDGFKYFAHNEDLGNEEGEISYSGIL-- 2060
Qy 595 HMWNGKTRNIYIPTASADNRGSLITINHEAGYNSYLFINGDEKVVSGYKFSVSDGQF 654
Db 2061 -----NFNKKIYY---FDDSFYAVGVWKDLEDGSKYYF---DEDTAEAYIGLSLINDQY 2109
Qy 655 WKERDVYDTRARKPEQFG--VPVTTLVGYDPEGTLSYIYPAMYGAFTYSDSDSNLS 713
Db 2110 YFNDDGI-----MQVGFVINDKYFYFSDSGIIES-----GVQNDIDNWFYI 2151
Qy 714 DNDQQLQV---DTKEGQLRFLANHRANNTVMKFNHNVPTESOPTOATLVCNKILDTK 770
Db 2152 DDNGIGVQIGVFDTSQDKYKYPAPA-----NTVNDNIY-----GQAVEYS 2189
Qy 771 SLTPAPE-----GLTYTVNGQALPAKNEGCIYVNSGKRYCPLPVGORS--GYSLPWNIV 823
Db 2190 GLVRVGEDVYFYGETTYTETGIWYDMENE-----SDKYFFNFPETKKACKGINLIDDI- 2241
Qy 824 GQEVYVDSGAKAKVLLSDWDLNLSY--NRIGEF--VGNVNPAD 861
Db 2242 --KYIFDEKGMRTGLISFENNNTYFNENGEMQFYGINIED 2280
```

RESULT 13

A83412

hypothetical protein PA1874 [Imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: A83412
R.Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoquchi, S.D.; Warrenner, P.; Hickey, M.J.; B.
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: A83412
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2468 <STO>
A:Cross-references: GB:AE004613; GB:AE004091; NID:99947856; PIDN:AAG05263.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA1874

Query Match 3.2%; Score 148.5; DB 2; Length 2468;
Best Local Similarity 19.5%; Pred. No. 0.92;
Matches 182; Conservative 90; Mismatches 301; Indels 361; Gaps 45;

QY 33 NTSOPINDLQGSLLAEVKEFAQSQILP-----AHP-----KEGDSOPH----- 69
DB 324 NNSTPTVTEAPOTTAPATDQVQVAPDGSSTVTCNAEPGATGVGDTGDCQPTTVVVGPG 383
QY 70 -----LTSRLSKLLLRPVKADKTPVOVEARD-----DNKKIL----- 103
DB 384 GSFEVPLNPPLTNGETVTIVITDPAGNSSTPVTAEAPDPDPAPQVNASGSLVSGTAEAG 443
QY 104 -----GTLTLYPPSSLPD-----TIYHLQGV 124
DB 444 VTIVITDGNPNIGOTSADANGWMSPTGSQLPDGTVVNVVARDAGNSSPATSIIVDGV 503
QY 125 PEGGIDFTPHNGPKKIINTVAENK---LSDASGSIHSHLTNNALVEIHTANGVRVDI 181
DB 504 APNAPVPSNSE---LSGTAEPGSSVTLTDGNGNPGICQTAD-----ANGNW---S 550
QY 182 YLPQGDLEGGKVRVFS---SAGYSS-----TFYGDVKVTLVSGNTLLFKFYVNGQWF 231
DB 551 FTPSTPLPDGTVVNVVARDAGNSSPASVTVDAVAPATPTVDPSSGNTTL----- 600
QY 232 RSGELENNRITYAQHITWSAELPAHWIVPGNLVVIKGNLSGRINDIKIGAPCELLHTID 291
DB 601 -SGTAE-----PGSSVTLTDGN-----GNP----- 619
QY 292 IGLMTTPDRDFPAKDEAHREYFOTIPVSRMIVNNAAPLHLKEVMLPTGELLTD----- 346
DB 620 IGVQTA-----DGSQNWTFPTPLPNGTVVATATD-----PSGNASSPASVIV 664
QY 347 -----MDPONGGWHSGTHQRIGKELVSHGIDNANYGLNSTAGLGNSHPYVVAQL 397
DB 665 DAVAPATPVVNPSSGNTLSGT-----ABPGATVTLTDG-NGNP--IGQV 705
QY 398 AAHNSRGNTA-----NCIQVH-----GGSGGGIVTLDS-----TL 428
DB 706 TAGDS-GNMSFTPTPLPNGTVVNVATATDASGNTSAGSSVTVDSVAPATPVINPNSGNTIL 764
QY 429 GNEFSHEVGHVGLGHYVDFKGSVHRSANNNSTWGDGDKRRFIPNFYPS-OTNEKSC 487
DB 765 SG--TAEPGSSVTL---TDGNGNPIQVQTAGDGSNW-----STPTPLADGTV 808
QY 488 LNNQCOPEPDGKHGFDAMAGSGFSAANRFTMYTPNSSAIIORFFENKAVFDSRSSTGF 547
DB 809 VNATATDP-----AGNTSGGSGTVDGVAPTPTV-----NLSNGSSLSGTA- 850
QY 548 SKWNADTQEMPEYHTIDRAEQITASVNELSEKMAELMAEYAVVVMHNGNHWTRNIYI 607
DB 851 -----EPGS-----TVILTDGNGNPIAEVTAD-----GSGNWT---YT 880
QY 608 PTASADNRGSILITINHEAGVNSLYFINGDEKVVSOQYKKSFSVDGQFWKERDVEDVTREAR 667
DB 881 PSTPIANGTVVNVVAQDAAGNS-----SPGASVT-----VDSQAPA 916
QY 668 KPQFGVPVTLVGYDPEGLTSSYIYPAMYGYFTYSDDSQLNSDNCQLQVDTKEGQ 727

DB 917 APVVNPSNGTTLSGTAEPGATV-----TLTDGNGNPIG---QVTAD-GSGN 958
QY 728 LRFLRANHRANNVYKPHINVPTESOPTQATLVCKNNKILDTKSLTAPAGLTYTVNCOA 787
DB 959 WSFTFGTPLANGTVVN-----ATASDPTGNTSPASTTVD--SVAPAAP-VVNPSPNGAE 1009
QY 788 LPAKENECCIVSVNSGKRYCLFVPGQSGYSLPDW 821
DB 1010 ISGTAEPGATVTLTDGSGN--PIGQVTADGSGNW 1041

RESULT 14
I40805
collagenase - Clostridium histolyticum
C:Species: Clostridium histolyticum
C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 21-Jul-2000
C:Accession: I40805; T44355
R:Yoshihara, K.; Matsushita, O.; Minami, J.; Okabe, A.
J. Bacteriol. 176, 6489-6496, 1994
A:Title: Cloning and nucleotide sequence analysis of the colH gene from Clostridium
A:Reference number: I40805; MUID:95050206; PMID:7961400
A:Accession: I40805
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1021 <RES>
A:Cross-references: GB:D29981; NID:9563954; PIDN:BAA06251.1; PID:g710023
R:Matsushita, O.; Jung, C.M.; Katayama, S.; Minami, J.; Takahashi, Y.; Okabe, A.
J. Bacteriol. 181, 923-933, 1999
A:Title: Gene duplication and multiplicity of collagenases in Clostridium histolytic
A:Reference number: Z27252; MUID:99121032; PMID:9922257
A:Accession: T44355
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1021 <MAT>
A:Cross-references: EMBL:AB014075; NID:93868863; PIDN:BAA34542.1; PID:g3892646
A:Experimental source: strain JCM 1403
C:Genetics:
A:Gene: colH
C:Superfamily: microbial collagenase

Query Match 3.1%; Score 148; DB 2; Length 1021;
Best Local Similarity 18.6%; Pred. No. 0.25;
Matches 169; Conservative 121; Mismatches 330; Indels 288; Gaps 46;

QY 97 DNNKILGTLYPPSSLPDTIYHLDGVPGEIGDTFPHNGTKKIINTVAENKLSASGS 156
DB 263 DNNNSWI-----IDNGIYHI--APLGKL-----HSNNKIGITLTVMK----- 299
QY 157 SIHSHLTNNALVEIHTANGRVYRDIYLPQGPDLGKQV---RFVSSAGYS---STVFYGD 210
DB 300 -VVPYLSMQ-----HLOSADQIKRHY--DSKDAEGNKIPLDKPKGKGYCPTKTYTDD 351
QY 211 RYVTLVSGVNTL-----LFKYVNGQWFRSGELENRRITYAQHINSAELPAHWIVPG 260
DB 352 GKVIIRKAGARVEEEKYKRLYKASKEYNSQFFR-----VYG 386
QY 261 LNLVIKOGNLSGRINDIKIGAPCELLHTI-----DIGMLTTPDRDFPAKDEAHRE 313
DB 387 IKPLEEGPDDILLNVIYNSPEEYKLSVLYGDTNNGMTEPEGTF-FYIERAQES 445
QY 314 YFOTIPVSRMIVNY-----APLHLKEVMLPTGELLTMDPONGGWHSGTMRQRIKEL 367
DB 446 TYTLEELFRHEYTHYLOGRYAVPGWGRTKLYDNDRLTWEEG-----GAL 492
QY 368 VSHGIDNANYGLNSTAGLGNSHPYVVAOLAHSNGN---YANGIQVHGGSGGGIVTL 425
DB 493 FA-----GSTRSGILPRKS--IVSNI--HNTTRNNRYKLSDTVH----- 528
QY 426 STLGNEFSHEVGHVGLGHYVD-----GFGKSVHRSANNNSTWGDGDKRRFIPNRY 478
DB 529 SKYGASFEF---YNTAC-MFMDYTNKMDGILNKLNDLAKNNDVD-GYDN-----Y 574

Db	715	VLADYAEV	DDNANKYV	NKSGBOV	GNIDW	EY----	ANQSG--	STVS	NAT----	VTDTLG	766
Qy	146	EVNKLS	DAGSS	THSHLT	NNALVE	HTANG	RWVRD	LYL	POGP-----	DLECK	192
Db	767	TGOKL	-DT	SIKYK	KSOT-----	SVTKGM	QOESN	MPIS	POEYDL	KTGVDEES	817
Qy	193	MVRFS	SAGSY	STVFY-----	GDRKVT	LSVGN	TLLFKY	VNGN	QFRSGE	LENNR	245
Db	818	QVKFK	NEINQ	SVYIK	QTALT	TSDET	TAQ	IGNSVTF--	TGDNIT	KGETEK	869
Qy	246	HMSAE	LPAMW	VPGLN	LVKQGN	LSRLND	IKIGAP	CELLHT	IDIGML	TTPRO--	303
Db	870	-----	NI	EKVIT	TDG	TG-----	TG	ETKIL	NKVDK	APIS	909
Qy	304	-FARK	KEAREY	FQ----	TIPVSR	MVNNY	APLHL	KEVLM	PTGELL	DDMPGN	359
Db	910	LYANDE	KVDQT	TDKNC	VEFDD	LV	GDYT--	LKEVSA	PGVTL	TASTEN-----	959
Qy	360	RQRIG	KELVSH	GDN	NANG	UNSTAG	LENSH	PVYV	QAALAH	NRSGV	415
Db	960	QVKLE	QDEK	VVQV	MNEK	MPIKET--	GE-----	VHLVKT	DKATG	ATLAGAE	1005
Qy	416	---SGG	---GG	IVTLD	STLGN	ESHEV	GHV	GHVY-----	DGFK-----	-----	450
Db	1006	YKSGA	ELO	NGL	TT-----	DENGEL	TIH	LDLGS	YLYKET	KAPG	1059
Qy	451	---GS	HRSA	ENNN	TW	GD	KRFIP	NP	PSQ	TNEK-----	490
Db	1060	GOVDA	IEIOA	EN-----	EKDL	GEAV	LTKVD	SETNA	KLSGAK	FNLLDS	1112
Qy	491	-----	QCQEP	FDGH	KFGF	DAMAG	SP	SAANR--	FTWYTP	NSSA	539
Db	1113	DERGE	IRVQ	NLEPGD	---YAFQ	ETEAP	TNYDL	ATNT	WPTF	IVAGOT	1166
Qy	540	DSRS	TGFS	KWNA	DQ	EMEP	TEHT	IDRAE	QITAS	VNELSE	599
Db	1167	---TG-	RPD	VD	TG	EVIL	VKQDS	ATG	LEG	AVFDL-----	1209
Qy	600	NWTR	NIYIP	TASAD	NRG	SILTI	NHEAG	YN	SYLF	INGDEK	659
Db	1210	-----	TTD	ANGE	ITV	NLAP	KY	SF----	KETKA	PEGYELA-----	1243
Qy	660	VVD--	TREAR	KPEQ	GPV	PV--	TLG	YD	PECT	LSSYI	711
Db	1344	VWEET	IAPNO	PEKIT	ITIA	ENK	LAPID	PAG	SVK-----	-----	1290
Qy	712	-----	LSDN	QCLO	VD	TK-----	EGOLRF	-RLA--	NHRA	NTVM	1290
Db	1291	AEFSL	IAENG	ETL	QNL	KTD	DEAGE	LEV	NNLAP	GNRIQ	1347
Qy	761	VCNNK	ILDT	KS	LTPA	PEGL	YTV	NGOAL	PAKENE	GCIV	817
Db	1348	VAN----	DT	SO	VT	VAIE-----	NAKLEP	DV	AE	TGAV	1393
Qy	818	LPD	IVQ	GEY	TV	D	S	GAKAY	LLSD	WNLS	874
Db	1394	LLD-----	ESG----	KV	IOA--	NLT	T	D	ENGE	IF	1437
Qy	875	FSKPR	SRV	V	884						
Db	1438	AEQ	PWN	FQ	1447						

Search completed: October 8, 2003, 16:28:15
Job time : 27.7788 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 8, 2003, 16:10:42 ; Search time 54.9419 Seconds
(without alignments)
2559.645 Million cell updates

Title: US-10-002-309B-2

Perfect score:

Sequence: 1 MKLKYLSTILAPLAIGVFS.....AWNGQYLDFSKPRSMRVVYK 886

Scoring table: BLOSUM62

scoring cable: BLOSOM2
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_19Jun03:*

1:	/SIDSL/cgcdata/geneseq/geneseqp-emb1/AA1980.DAT:*
2:	/SIDSL/cgcdata/geneseq/geneseqp-emb1/AA1981.DAT:*
3:	/SIDSL/cgcdata/geneseq/geneseqp-emb1/AA1982.DAT:*
4:	/SIDSL/cgcdata/geneseq/geneseqp-emb1/AA1983.DAT:*
5:	/SIDSL/cgcdata/geneseq/geneseqp-emb1/AA1984.DAT:*
6:	/SIDSL/cgcdata/geneseq/geneseqp-emb1/AA1985.DAT:*
7:	/SIDSL/cgcdata/geneseq/geneseqp-emb1/AA1986.DAT:*
8:	/SIDSL/cgcdata/geneseq/geneseqp-emb1/AA1987.DAT:*
9:	/SIDSL/cgcdata/geneseq/geneseqp-emb1/AA1988.DAT:*
10:	/SIDSL/cgcdata/geneseq/geneseqp-emb1/AA1989.DAT:*
11:	/SIDSL/cgcdata/geneseq/geneseqp-emb1/AA1990.DAT:*
12:	/SIDSL/cgcdata/geneseq/geneseqp-emb1/AA1991.DAT:*
13:	/SIDSL/cgcdata/geneseq/geneseqp-emb1/AA1992.DAT:*
14:	/SIDSL/cgcdata/geneseq/geneseqp-emb1/AA1993.DAT:*
15:	/SIDSL/cgcdata/geneseq/geneseqp-emb1/AA1994.DAT:*
16:	/SIDSL/cgcdata/geneseq/geneseqp-emb1/AA1995.DAT:*
17:	/SIDSL/cgcdata/geneseq/geneseqp-emb1/AA1996.DAT:*
18:	/SIDSL/cgcdata/geneseq/geneseqp-emb1/AA1997.DAT:*
19:	/SIDSL/cgcdata/geneseq/geneseqp-emb1/AA1998.DAT:*
20:	/SIDSL/cgcdata/geneseq/geneseqp-emb1/AA1999.DAT:*
21:	/SIDSL/cgcdata/geneseq/geneseqp-emb1/AA2000.DAT:*
22:	/SIDSL/cgcdata/geneseq/geneseqp-emb1/AA2001.DAT:*
23:	/SIDSL/cgcdata/geneseq/geneseqp-emb1/AA2002.DAT:*
24:	/SIDSL/cgcdata/geneseq/geneseqp-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4709	100.0	886	23	Escherichia coli s
2	153.5	3.3	2366	17	C. difficile toxin
3	153.5	3.3	2366	19	Clostridium diffic
4	149	3.2	1249	23	Alpha-isomaltosylg
5	149	3.2	1284	23	Alpha-isomaltosylg
6	149	3.2	1284	24	Bacillus globispor
7	149	3.2	1284	24	Protein #1 related
8	147.5	3.1	1530	23	Listeria monocytog
9	145	3.1	929	23	Alpha-isomaltosylg

10	145	3.1	965	23	ABG30565	Alpha-isomaltosylg
11	144.5	3.1	2057	21	AB100667	L. mesenteroides a
12	144	3.1	1286	23	ABG30564	Alpha-isomaltosylg
13	144	3.1	2314	22	ABG69136	M. catarrhalis les
14	143.5	3.0	1251	23	ABG30538	Alpha-isomaltosylg
15	142	3.0	784	24	ABP715687	Lactococcus amylo
16	141.5	3.0	1992	17	AAW04505	Moraxella 200 kDa
17	141.5	3.0	1992	22	AAW69133	M. catarrhalis str
18	141.5	3.0	1992	22	AAW69137	M. catarrhalis M56
19	141.5	3.0	2047	22	AAW69134	M. catarrhalis str
20	141	3.0	2053	22	AAW69135	M. catarrhalis str
21	139	3.0	1041	18	AAW11866	ATM1 gene product,
22	139	3.0	1041	20	AAV29830	Saccharomyces cere
23	139	3.0	1041	20	AAV16788	S. cerevisiae acid
24	138.5	2.9	973	21	AAV196274	Chlamydia POMP91B
25	138.5	2.9	992	23	ABB90527	Chlamydia pneumonia
26	134.5	2.9	992	21	ABAB1843	Haemophilus influe
27	134.5	2.9	998	21	ABAB1842	Haemophilus influe
28	134.5	2.9	1026	20	AAW94678	Bacteriophage T4 t
29	134.5	2.9	1026	20	AAW92358	Bacteriophage T4 t
30	134.5	2.9	1026	22	AAW35354	Bacteriophage T4 g
31	134	2.8	1781	23	AAU74519	Lactobacillus reut
32	132.5	2.8	2139	24	ABP71294	M. catarrhalis sur
33	132	2.8	815	22	AAU33610	Pseudomonas aerugi
34	132	2.8	1413	18	AAW20725	H. pylori secreted
35	131.5	2.8	1026	17	AAW97374	Phage T4 tail fibr
36	131	2.8	576	24	ABU26400	Aspergillus fumiga
37	131	2.8	576	24	ABU26526	Aspergillus fumiga
38	130	2.8	684	23	ABG73040	Cryptococcus nodae
39	128	2.7	1382	19	AAV11001	H. pylori ORF 11e
40	127.5	2.7	918	21	AAV94672	Haemophilus antige
41	127.5	2.7	2188	22	ABW65784	Drosophila melanog
42	127	2.7	711	19	AAW73025	Helicobacter pylori
43	127	2.7	1612	19	AAW65088	R. prowazekii S-la
44	126.5	2.7	2383	21	AAAB15945	E. coli proliferat
45	126	2.7	1978	20	AAV27230	Amino acid sequen

ALIGNMENTS

RESULT 1	
AAE24015	
ID AAE24015 standard; Protein; 886 AA.	
XX	
AC AAE24015;	
XX	
AC AC	
XX	
DT 23-SEP-2002 (first entry)	
DE	
XX Escherichia coli stcE protein encoded by pO157 plasmid DNA.	
DE	
XX pO157 plasmid; stcE protein; haemolytic uraemic syndrome; p	
KW C1-esterase inhibitor; enterohaemorrhagic pathogen; antitoxin	
KW colitis; antibacterial; anti diarrhoeic.	
XX	
OS Escherichia coli O157:H7.	
XX	
FH Key Location/Qualifiers	
FT Peptide 1..23	
FT /label= Signal_peptide	
FT Protein 24..886	
FT /note= "Mature human secreted protein"	
FT Binding-site 434..444	
FT /note= "Metalloprotease binding domain"	
XX	
XX WC200234918-A2.	
XX	
PN PN	
XX	
PD PD	
XX	
PF 26-OCT-2001; 2001WO-US47719.	
XX	
PR 26-OCT-2000; 2000US-243675P.	
XX	

10/26/97 10/26/97 10/26/97

xx PA (WISC) WISCONSIN ALUMNI RES FOUND.
 xx DR Welch RA, Latham WH;
 xx PI N-PSDB; AAD38751.
 xx WPI; 2002-471441/50.
 xx DR N-PSDB; AAD38751.
 xx PT New p0157 plasmid-specified polypeptide found in Escherichia coli and
 xx PT other enterohemorrhagic Escherichia coli, that binds to and cleaves
 xx PT Cl-esterase inhibitor, useful for diagnosing and treating colitis -
 xx PS Claim 1; Page 52-56; 58pp; English.
 xx CC The present invention relates to novel p0157 plasmid-specified proteins
 xx CC found in Escherichia coli EDL933 and other enterohemorrhagic E. coli,
 xx CC designated StcE, that bind to and cleave Cl-esterase inhibitor. Sequences
 xx CC of the invention are useful for diagnosing, preventing or treating
 xx CC haemolytic uraemic syndrome or colitis in a subject infected with an
 xx CC enterohemorrhagic pathogen expressing inhibitor protein. They are
 xx CC useful for testing a molecule for the ability to reduce proteolysis
 xx CC of Cl esterase inhibitor by inhibitor protein. The present sequence
 xx CC is E. coli O157:H7 StcE protein encoded by p0157 plasmid DNA.
 xx SQ Sequence 886 AA;
 Query Match 100.0%; Score 4709; DB 23; Length 886;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 886; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MKLYLSCTILAPLAIGVFSATAADNNSAIVNTSOPINDLOGSLAAEVEKFAQSOILPAH 60
 Db 1 MKLYLSCTILAPLAIGVFSATAADNNSAIVNTSOPINDLOGSLAAEVEKFAQSOILPAH 60
 QY 61 PREGSQPHLSRLSKLLVPRVADKTPVQVEARDNNKILGTLTLYPPSSLPDTIYH 120
 Db 61 PREGSQPHLSRLSKLLVPRVADKTPVQVEARDNNKILGTLTLYPPSSLPDTIYH 120
 QY 121 LDGVPPEGIDFTPHNGTKKIIINTVAEVNKLSDAGSSISHLTNNALVEIHTANGRWYRD 180
 Db 121 LDGVPPEGIDFTPHNGTKKIIINTVAEVNKLSDAGSSISHLTNNALVEIHTANGRWYRD 180
 QY 181 IYLPQGPDLGKRWVFSAGSYSTVFYGDARKVTLVSGNTLLFKYVNGQWFRSGELENNR 240
 Db 181 IYLPQGPDLGKRWVFSAGSYSTVFYGDARKVTLVSGNTLLFKYVNGQWFRSGELENNR 240
 QY 241 IYTAQHISAEPLAHWIVPGLNLVTKQNLGRLNDIKIGAPGELLHTIDIGMLTTPRD 300
 Db 241 IYTAQHISAEPLAHWIVPGLNLVTKQNLGRLNDIKIGAPGELLHTIDIGMLTTPRD 300
 QY 301 RFDFAKDKEAHREYFOTIPVSRMIVNNYAPLHLKEVMLPTGELLTDMDPGNGWHSGTMR 360
 Db 301 RFDFAKDKEAHREYFOTIPVSRMIVNNYAPLHLKEVMLPTGELLTDMDPGNGWHSGTMR 360
 QY 361 QRIGKELVSHGIDNANYGLNSTAGLGENSHPYVAQLAHNSRGNANGIQVHSGSGGG 420
 Db 361 QRIGKELVSHGIDNANYGLNSTAGLGENSHPYVAQLAHNSRGNANGIQVHSGSGGG 420
 QY 421 IVTLDTLGLNEFSHEVGHNYGLGHVYDGFKGSVHRSANNSTWGDGKRFPINFPYS 480
 Db 421 IVTLDTLGLNEFSHEVGHNYGLGHVYDGFKGSVHRSANNSTWGDGKRFPINFPYS 480
 QY 481 QNTEKSLNQCQOEPDGHKFGFDAMAGSPESAANRFTMTYTPNSSAIIQRFENKAYFD 540
 Db 481 QNTEKSLNQCQOEPDGHKFGFDAMAGSPESAANRFTMTYTPNSSAIIQRFENKAYFD 540
 QY 541 SRSSGFGSKWADTQEMEPYEHTIDRAEQITASVNELSEKMAELMAEYAVYKVDWNGN 600
 Db 541 SRSSGFGSKWADTQEMEPYEHTIDRAEQITASVNELSEKMAELMAEYAVYKVDWNGN 600
 QY 601 WTRNIYIPTASADNRGSIILTINHEAGYNSYLPINGDEKVSQYKKSFSVDSQFQWKRVDV 660
 Db 601 WTRNIYIPTASADNRGSIILTINHEAGYNSYLPINGDEKVSQYKKSFSVDSQFQWKRVDV 660

QY 661 VDTREARKPEQFGVPVTTLVGYDPEGTLSSTYYPAMYGAIFYTSDSNDQCLQ 720
 Db 661 VDTREARKPEQFGVPVTTLVGYDPEGTLSSTYYPAMYGAIFYTSDSNDQCLQ 720
 QY 721 VDTKEGQLRFLANHRANNTVNNKFHINVPTESQPTQATLVCNNKILDTKSLTPAPEGLT 780
 Db 721 VDTKEGQLRFLANHRANNTVNNKFHINVPTESQPTQATLVCNNKILDTKSLTPAPEGLT 780
 QY 781 YTVNGQALPAKENGCGICIVSVNSGRKRYCLPVQBSGYSLPDWIVGQVYVDSGAKAKVLLS 840
 Db 781 YTVNGQALPAKENGCGICIVSVNSGRKRYCLPVQBSGYSLPDWIVGQVYVDSGAKAKVLLS 840
 QY 841 DWDNLSYNRIGEGFVGNVPADMKKVKAWNGOYLDFSKPRSMRVYK 886
 Db 841 DWDNLSYNRIGEGFVGNVPADMKKVKAWNGOYLDFSKPRSMRVYK 886
 RESULT 2
 AAR95011
 ID AAR95011 standard; Protein; 2366 AA.
 XX AAR95011;
 AC AAR95011;
 XX 08-JUL-1996 (first entry)
 DT C. difficile toxin B.
 DE C. difficile toxin B.
 XX Toxin B; cytotoxin; enterotoxin; fusion protein; antitoxin;
 KW diarrhoea; therapy; diagnosis; vaccine.
 XX Clostridium difficile VPI strain 10463 (ATCC 10463).
 XX WO9612802-A1.
 PD 02-MAY-1996.
 XX 23-OCT-1995; 95WO-US13737.
 PF 07-JUN-1995; 95US-0480604.
 PR 24-OCT-1994; 94DS-0329154.
 PR 16-MAR-1995; 95US-0405496.
 PR 14-APR-1995; 95US-0422711.
 XX (OPHI-) OPHIDIAN PHARM INC.
 XX Firca JR, Kink JA, Padhye NV, Stafford DC, Thalley BS;
 PI Williams JA;
 XX WPI; 1996-230603/23.
 DR N-PSDB; AAT29247.
 XX Fusion proteins comprising non-toxin protein and part of toxin -
 PT useful to form anti-toxins against Clostridium botulinum type A, and
 PT C. difficile type toxins, and to treat C. difficile intoxication,
 PT partic. diarrhoea
 XX Claim 36; Page 313-323; 434pp; English.
 PS Clostridium difficile VPI strain 10463 toxin B (AAR95011), the
 CC product of the toxin B gene (AAT29247), is a cytotoxin associated
 CC with diarrhoeic disease. It can be obtd. by expression in
 CC transformed E. coli hosts of portions of DNA that together cover the
 CC entire toxin B gene. Toxin B, and portions of it (see also AAR95012-13,
 CC AAR75371-72 and AAR95018), pref. expressed as fusions to polypeptide
 CC affinity tags or maltose binding protein, are used to raise avian
 CC antibodies useful as antitoxins or diagnostics, and in vaccine prodn.
 XX Sequence 2366 AA;
 SQ
 Query Match 3.3%; Score 153.5; DB 17; Length 2366;
 Best Local Similarity 18.6%; Pred. No. 0.033;
 Matches 186; Conservative 155; Mismatches 321; Indels 339; Gaps 51;

Qy	27	NSAIYFNTSOP	INDLQGS	LAAEYFK	QAQO	ILPAHP	KEG	SQPHLT	SLRKS	LLLPV	KAD	86
Db	1453	NSELOKNIPYS	FVDSECK	-----	ENG	FING	STEG	----	LFV	SEL	PDV	1500
Qy	87	DKTP	-----	VQ	EAR	DDNN	KILG	-----	TL	LYP	SSLP	122
Db	1501	DSKPSFG	YSGN	NLKD	VKIT	QDN	NIL	TGYL	KDD	KIS	LSLT	1560
Qy	123	GVPEGG	IDFP	PHNG	TKKI	INT	VAE	NKLS	DAS	GSS	THSL	180
Db	1561	GVAE	-----	IL	KPM	RKGN	TN	TS	DL	MS	FLES	1609
Qy	181	-----	IY	PG	DP	LE	KMYR	FV	SS	AG	YS	216
Db	1610	FIISG	TTSIG	EF	IC	DEN	NI	QY	FI	KF	NTLE	1668
Qy	217	VGN	TLL	---	FR	YNG	QW	FR	SE	LEN	NR	260
Db	1669	ISST	YFN	SOK	LYG	-----	ID	SC	VN	KV	VI	1721
Qy	261	LN	LV	KOG	N	LS	GR	LD	KI	-----	GAP	312
Db	1722	ANT	INE	KIN	YN	-----	IN	D	SI	RY	VS	1779
Qy	313	EYF	-----	Q	T	P	S	R	M	V	N	338
Db	1780	LS	NF	S	D	K	O	V	P	S	E	1838
Qy	339	-----	PT	CELL	TD	-----	MD	P	G	N	G	379
Db	1839	N	D	S	L	Y	F	K	P	V	N	1888
Qy	380	N	S	T	A	G	L	G	E	N	S	431
Db	1889	N	Q	S	G	-----	V	L	T	G	V	1924
Qy	432	F	S	H	E	V	G	H	N	Y	G	488
Db	1925	F	T	G	L	I	D	E	N	I	Y	1961
Qy	489	N	N	O	C	E	P	D	H	K	G	542
Db	1962	-----	T	K	A	F	K	G	L	N	O	2002
Qy	543	S	--	S	T	G	F	S	K	-----	W	594
Db	2003	G	V	M	K	V	T	E	D	G	K	2060
Qy	595	H	M	N	G	N	T	R	I	Y	P	654
Db	2061	-----	N	F	N	K	I	Y	-----	F	D	2109
Qy	655	W	K	E	R	D	V	D	T	R	E	713
Db	2110	Y	F	N	D	D	G	I	-----	M	O	2151
Qy	714	D	N	O	C	O	L	O	V	-----	D	770
Db	2152	D	D	N	I	V	O	I	G	V	F	2189
Qy	771	S	L	T	P	A	P	E	-----	G	L	823
Db	2190	G	L	V	R	G	E	D	V	Y	F	2241
Qy	824	G	O	E	Y	V	D	S	G	A	K	861
Db	2242	--	K	Y	F	D	E	K	G	I	M	2280

RESULT 3
AAW68388
ID AAW68388 standard; protein: 2366 AA.

РЕСПУБЛИКА

RESULT 3
111369311

AAW68388
ID AAW68388 standard; Protein; 2366 AA.

[illegible]

Db 1722 ANYINEXINVN--INDISIRYVNSDGNDFILMSTSEENKVSVQVIRFVNVFKDKTLANK 1779
 QY 313 EYF-----QTIPYSRMIVNNAVLHLKEVML----- 338
 Db 1780 LSNFSDKQDVPVSEIL-SFTPSYYEDGLIGYDLGLVSLYNEKFFYNNGFMVSGLIYI 1838
 QY 339 -----PTGELLTD-----MDPENGCGHSGTMRQIRIGKELVSHGIDNANGL 379
 Db 1839 NDSLYFKPPVNNLITGVTVGDDKYFNPINGCAAS-----IGETI-----IDDKNYF 1888
 QY 380 NSTAGLGESHPPYVVAQLAAHNSRGN-----ANGIQVHGGSGGGIVTLDSTLGN----- 431
 Db 1889 NQSG-----VLGTGVFSTDEGKFKYFAPAN-----TIDENLEGEAID 1924
 QY 432 PSHEVGHNYGLGHVVDGFKGSHVRSANNNSTWGW---DGDKKRFIFNFPVPSQTNKSCL 488
 Db 1925 FTGLKIIDENIYFDDNYRGAVE-----WKELDGMHFVSP----- 1961
 QY 489 NNOCQEPFDGKHFGFDAMAGGSPFSAANRF--TWITPNSSAIORFF-----ENKAVTDSR 542
 Db 1962 -----TCKAFKGLNQIGDYKYFYFNSDGMOKGFVSINDNKHFFDSDS 2002
 QY 543 S--STGFSK-----WNADTQEMPEYEHITIDRAEQITASVNELSESKMAELMAEYAVVKV 594
 Db 2003 GVMKVGTEIDGKHFFAENGEMOIGVNTEDGFKYFAHNEDLGNEGEISYSGIL-- 2060
 QY 595 HMMNGNTRNIYIPTASADNRGSILTINHEAGYNSYLFINGDEKVSQGYKKFSVSDQGF 654
 Db 2061 -----NPNKKIY---PDSFTAVVGWKDLEDGSKYF---DEDTAEAYIGLSLINDGOY 2109
 QY 655 WKERDVDTREARKPEFG-VPVTTLVGYDPEGLTSSYIYPAMYAGFYTSDSONLS 713
 Db 2110 YFNDGI-----MOVGFYTIIDKVFYFSDSGIIES-----GVONIDDNYFYI 2151
 QY 714 DNDQLOLV---DTKEGQLRFLANHRANNTVMKFNHINVPESQPTQATLVCKNKKILDTK 770
 Db 2152 DONGIVQIGVETSDGYKYFAPA-----NTVNDNIY-----QCAVEYS 2189
 QY 771 SLTPEPE-----GLTYTVNGOALPAKENEGCIVSVNSGKRYCLPVGORS--GYSLPDMIV 823
 Db 2190 GLVRVGEDVYVFGETYTTIETGWIYDMENE-----SDKYFNPETKKACKGINLDDI- 2241
 QY 824 GOEYVVDGAKVLLSDNDLSY--NRIGEF-VGNVNPAD 861
 Db 2242 --KYFDEKIMRTGLISFENNNTYFENGEMQFYINIED 2280

RESULT 4
 ABG30537
 ID ABG30537 standard; Protein; 1249 AA.
 XX AC ABG30537;
 XX DT 07-OCT-2002 (first entry)
 XX DE Alpha-isomaltosylglucosaccharide synthase #1 mature protein.
 XX KW Alpha-isomaltosylglucosaccharide synthase; sugar production;
 KW cyclic tetrasaccharide; cyclo(-6)-alpha-D-glucopyranosyl(1-3);
 KW alpha-isomaltosyltransferase; viscosity modifier; sweetener;
 KW heat stabiliser; hydrolysis stabiliser; colour stabiliser; foodstuff;
 KW cosmetic; drug composition; enzyme.
 XX OS Unidentified.
 XX PN W020025708-A1.
 XX PD 18-JUL-2002.
 XX PF 09-JAN-2002; 2002WO-JP00052.
 XX PR 12-JAN-2001; 2001JP-0005441.
 XX

PA (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.
 XX Kubota M, Maruta K, Yamamoto T, Fukuda S;
 PI WPI: 2002-520129/55.
 DR N-PSDB; ABK88154.
 XX New alpha-isomaltosylglucosaccharide synthase of bacterial origin for
 PT the production of cyclic tetrasaccharide gum -
 XX Claim 1; Page 101-105; 144pp; Japanese.
 XX The invention describes novel microbial polypeptides having
 CC alpha-isomaltosylglucosaccharide synthase activity. The proteins are
 CC useful for producing a sugar (I) having at least three glucose units
 CC connected by alpha-1,4 bonds except for alpha-1,6 at the non-reducing
 CC end. The invention also describes a method for the production of the
 CC cyclic tetrasaccharide cyclo(-6)-alpha-D-glucopyranosyl(1-3)-alpha-D-
 CC glucopyranosyl(1-6)-alpha-D-glucopyranosyl(1-3)-alpha-D-glucopyranosyl
 CC (I) by treatment of (I) with alpha-isomaltosyltransferase.
 CC (II) and similar sugars in crystalline or syrup form are used as
 CC a viscosity modifier, sweetener, heat stabiliser, hydrolysis stabiliser
 CC and colour stabiliser for foodstuffs, cosmetics and drug compositions.
 CC This is the amino acid sequence of an alpha-isomaltosylglucosaccharide
 CC synthase mature protein.
 XX SQ Sequence 1249 AA;
 Query Match 3.24; Score 149; DB 23; Length 1249;
 Best Local Similarity 18.88; Pred. No. 0.027;
 Matches 170; Conservative 105; Mismatches 321; Indels 308; Gaps 42;
 QY 105 TLTLYPPSSLPDIYHLDGVPEG--GIDFTPHNGTKIINTVAEVNKLSDAGSSISHH- 161
 Db 19 TLTVDNGAEPDSDLIVQAVQNGILKVDYRPNSTPSAKTPMLDPNKTWSAVGATINTTA 78
 QY 162 -----LTNNALVEI-----HTANGRWROIYLPQGDPLEGKRVFSSAGYSTVF 207
 Db 79 NPMTITTSNNKIEITKNPVRMTVKKADGTTL--FWPESGGGVTFSDGVRFLLHATGDN---M 133
 QY 208 YGRKVTLSVGNITLLFKYVNGQWFRSGELENNRITVQAHIWSAELPAHVPGLNLVIKQ 267
 Db 134 YGIRSF-----NAFDSGGDLLRNSNHAH----- 158
 QY 268 GNLSGRINDIKIGAPGELLHTIDIGMLTTPROFDFAKDEAHREYFQ--TIPVSRMIV 325
 Db 159 ---AGEQGD---SGGPLINWSTAGYGLLVSDGYPYTDSTTGQMEFYVGGTPEGRRYA 211
 QY 326 NNTAPLHL-----KEYMLPTGELLTMDPONGGWHSGTMR-----Q 361
 Db 212 KQVEYYIMLGTPKREIMTDVGE-ITGKPPMLPKHSLGFNFMFEMDTNTEFTNNVDTYRAK 270
 QY 362 RIGKELVSHGIDNANYGLNSTAGLGESHPPYVVAQLAAHNSRGNVANGIQVHG----- 414
 Db 271 NIPIDAYAFDYDMKKYGETNYGEFAWNTTNPFSASTTSLKSTMD-ARGIKMIGITKPRIV 329
 QY 415 -GSGGGGIVT--LDSTLGNFSEHVGHNYGLGHVVDGFKGSHVRSAE---NNNSTWQWDG 468
 Db 330 TKDASANVTTCGTATNGGYF--YPGHN---EYQDYFIPVTYVRSIDIPYNANERAWFNH 383
 QY 469 DK---KRFIPNFYPSQTNKS-----CLNNOCQEPFDGKHFGFDAMAGGSP--- 511
 Db 384 STDALNKGIVGWNDETDKVSSGALYWFCNFTTGHMSQTMVEGGR---AYTSGAQRVW 439
 QY 512 -----FSAANRF--TWYTP-----NSSAIIQFFENKAVFDSRSSTGSKWN 551
 Db 440 QTARTFYPGAQRVATTLWSGDIGIYQNGKGERINWAAGHQ---EORAYMLSSVNGQVYK 496
 QY 552 ADT-----QEMEPY-----EHTIDRA 567
 Db 497 MDTGGFNOQDGTNNPNPDLYARMQFSALTPTVFRVHGNHQQRPWYFGSTAEASKEA 556
 QY 568 EQI-----TASVNELSESKMAELMAEY---AVVK--VDMWN-GNW----- 60-

Db	557	IQLRSLIPYMYAYERSAYENGCLVRPLMQAYPTDAAVKNYTDAMWFGDWLLAAPVDK	616
QY	602	---TRNIYIPTASADN---RGSILT-----INHEAGYNSYLFINGDEKQVWSQGYKSF	648
Db	617	QOTSNDIYLPSCSWIDYARGNAITGGOTIRYSVNPDTLTDMPLFI-----KKCA	665
QY	649	VSDGQFWERDQVDTREARKPEQGVPTVLVG--YYDPGTLSSYIYPAMYGAYGTYYS	706
Db	666	IIPTQ--KVQDYGQASVTSVDVDFPDTTQSSFTYYDDDG-----ASYNYESGTGF	715
QY	707	DDSNLSNDQCOLQVDTREGOLRFLANHRANNTVNNKPHINVTESQPTQATLVCCNKI	766
Db	716	KOMNTAQDNG-----SGLSLSTGLGAKSGSYTPALQSYI---VKLHGSAGTSVTNNS-	763
QY	767	LDTKSLTPEGLTYVNGQALPAKENECCIVTSVNSGKRYCLPVGORSYSLPDMIVGOE	826
Db	764	-----AAMTSYASLEAKKAAGEG-----WATGKD	788
QY	827	VYVD 830	
Db	789	IYGD 792	
RESULT 5			
ID	ABG30563		
AC	ABG30563 standard; Protein; 1284 AA.		
XX	ABG30563;		
DT	07-OCT-2002 (first entry)		
DE	Alpha-isomaltosylglucosaccharide synthase #1.		
KW	Alpha-isomaltosylglucosaccharide synthase: sugar production;		
KW	cyclic tetrasaccharide; cyclo(-6)-alpha-D-glucopyranosyl(1-3);		
KW	alpha-isomaltosyltransferase; viscosity modifier; sweetener;		
KW	heat stabiliser; hydrolysis stabiliser; colour stabiliser; foodstuff;		
KW	cosmetic; drug composition.		
OS	Unidentified.		
XX	WO200255708-A1.		
PN	18-JUL-2002.		
PD	09-JAN-2002; 2002WO-JP000052.		
PF	12-JAN-2001; 2001JP-0005441.		
PR	(HAYB) HAYASHIBARA SEIBUTSU KAGAKU.		
PA	Kubota M, Maruta K, Yamamoto T, Fukuda S;		
PI	WPI: 2002-520129/55.		
DR	N-PSDB; ABR88157.		
XX	New alpha-isomaltosylglucosaccharide synthase of bacterial origin for		
PT	the production of cyclic tetrasaccharide gum .		
XX	Disclosure; Page 120-126; 144pp; Japanese.		
CC	The invention describes novel microbial polypeptides having		
CC	alpha-isomaltosylglucosaccharide synthase activity. The proteins are		
CC	useful for producing a sugar (I) having at least three glucose units		
CC	connected by alpha-1,4 bonds except for alpha-1,6 at the non-reducing		
CC	end. The invention also describes a method for the production of the		
CC	cyclic tetrasaccharide cyclo(-6)-alpha-D-glucopyranosyl(1-3)-alpha-D-		
CC	glucopyranosyl(1-6)-alpha-D-glucopyranosyl(1-3)-alpha-D-glucopyranosyl		
CC	(1-1) (II) by treatment of (I) with alpha-isomaltosyltransferase.		
CC	(II) and similar sugars in crystalline or syrup form are used as		
CC	a viscosity modifier, sweetener, heat stabiliser, hydrolysis stabiliser		
CC	and colour stabiliser for foodstuffs, cosmetics and drug compositions.		

RESULT 6
 ABP70652
 ID ABP70652 standard; Protein; 1284 AA.
 XX AC
 XX ABP70652;
 XX DT
 XX 06-MAR-2003 (first entry)
 XX DE
 XX Bacillus globisporus protein encoded by SEQ ID 2.
 XX KW
 XX Cyclic tetrasaccharide; sweetening agent.
 XX OS
 XX Bacillus globisporus.
 XX FH Key Location/Qualifiers
 XX FT Peptide 1..35
 XX FT /label= Signal_peptide
 XX PN WO200272594-A1.
 XX PD 19-SEP-2002.
 XX PF 08-MAR-2002; 2002WO-JP02213.
 XX PR 09-MAR-2001; 2001JP-0067282.
 XX PA (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.
 XX PI Aga H, Higashiyama T, Watanabe H, Sonoda T, Kubota M;
 XX DR WPI; 2003-092811/08.
 XX DR N-PSDB; ABV99948.
 XX PT New cyclic tetrasaccharide compounds useful as sweetening agent for
 XX PT foods, drinks or pharmaceuticals -
 XX PS Disclosure; Page 122-129; 133pp; Japanese.
 XX CC The present invention relates to novel cyclic tetrasaccharide compounds:
 CC cyclo(-6-)-alpha-D-glucopyranosyl-(1-3)-alpha-D-glucopyranosyl-(1-6)-
 CC alpha-D-glucopyranosyl-(1-3)-alpha-D-glucopyranosyl-(1-) derivatives.
 CC The tetrasaccharide compounds are used as sweetening agents for foods,
 CC drinks or pharmaceuticals. The present protein was used to illustrate the
 CC invention.
 XX SQ Sequence 1284 AA:
 Query Match 3.2%; Score 149; DB 24; Length 1284;
 Best Local Similarity 18.8%; Pred. No. 0.028;
 Matches 170; Conservative 105; Mismatches 321; Indels 308; Gaps 42;
 QY 105 TLTLPPSSLPDIYHLGVPEG--GIDFTPHNGTKKIINTVAEVNKLSDASGSIHSH- 161
 DB 114 TLTVNGAEPDOLLIVQVQNGILKVDYRNSITPSAKTPLDPNKTWSAVGATINTTA 113
 QY 162 -----LTNALVEI-----HTANGWRVDIYLPQGPDLKGMVRFVSSAGYSTVF 207
 DB 114 NPMITTSNMKIEITKPNVMTVKKADGTTL--FWEPSSGGVFSGVFLHATGDN---M 168
 QY 208 YGDRKVTLSVGNLTLFKYVNGQWFSRSELENRTIYAQHIWSAELPAHWIYVGLNLVIKQ 267
 DB 169 YGIRSF-----NAFDSGGDLRLNSSNHAH----- 193
 QY 268 GNLSGLRLNDIKIGAPGELLHTIDIGMLTTPRDRDFAKDKEAHREYFQ--TIPVSRMIV 325
 DB 194 ---AGEQD---SGPLWSTAGVGLLVDSGGVPPYTDSTTGQEFYGGTTPGGRYA 246
 QY 326 NNYAPLHL-----KEVMLPTGELLTMDPGNGNHSGTWR-----Q 361
 DB 247 KQVEYIIMLGTPKPEIMTDVGE-ITGKPPMLPKWSLGFPMFEWDTNQTEFTNNVDITYRAK 305
 QY 362 RIGKELVSHGIDNANYGLNSTAGLGENSHPPYVAQLAAHNSRGNTYANGIQVHG----- 414

Db 306 NIPIDAYAFDYDMKKYGETNYGEFANNTNPPSASTTSLKSTMD-AKGIKMGITKPRIV 364
 QY 415 -GSGGGGIYV--LDSTLGNFEFSHEVGHNTGLHYVDGFGSGVHRSAE-----NNNSTWGWG 468
 Db 365 TKDASANVTTQGTDATNGGYF--YPGHN-----EYODYFIPVTYRSDPYNANERAFWNH 418
 QY 469 DK---KRFIPNFYPSOTNEKS-----CLNNQCOEPDFDGHKFGFDAMAGSP--- 511
 Db 419 STDALNKGIVGWNDETDKVSSGGALYNTGNTTGHMSOTNTEGGR-----ATSGQRVW 474
 QY 512 -----FSAANRF--TWYTP-----NSSAIIOEFFENKAVFDSRSSTGFSKNW 551
 Db 475 QTARTFPGAQRYATTILWSDIGIQYNGERINWAAQM-----EQRAVMLSSVNGQVKGW 531
 QY 552 ADT-----QEMEPY-----EHTIDRA 567
 Db 532 MDTGGFNQDGTNNPNPDLYARMQFSAITPVFRVHGNHQQORQPMWYFGSTAEEASKEA 591
 QY 568 EQI-----TASVNELSESKMAELMAEY---AVVK--VHMWN-GNW----- 601
 Db 592 IQLRSLIPYMYAYERSAYENGNGLYRPLMQAYPTDAVKNYTDAMWFGDWLLAAPVVK 651
 QY 602 ---TRNIYIPTASADN--RGSILT-----INHEAGYNSYLFINGDEKVVSGYKKSP 648
 Db 652 QQTSKDIYLPSSGWIDYARGNAITGGQTIRYSVNPOTLTDMPLEI-----KKA 700
 QY 649 VSDGQFWKRDVYDTRAREKPEQFGVPTVLVG--YDPEGTLSSYIYPAMYGAYGFTYS 706
 Db 701 IIPQ--KVODYGQASVTSVDVDFDTQSSFTYDDG-----ASYNYESGTYF 750
 QY 707 DSDNLSDNDCLQVDTKEGQLRFRLANHRANNTVMNKFHINVTESQPTQATLVNCKI 766
 Db 751 KONMTAQDNG-----SGSLFTLGAKSGSYTPALQSYI---VKLHGSAGTSVTNNS- 798
 QY 767 LDTKSLTPAPEGLTYTVNQALPAKENEGCIVSVNGSKRYCLPVGORSYSLPDWIVGOE 828
 Db 799 -----AAMTSYASLEAKAAAGE-----WATGKD 823
 QY 827 VYVD 830
 Db 824 IYGD 827
 RESULT 7
 ABP57724
 ID ABP57724 standard; Protein; 1284 AA.
 XX AC ABP57724;
 XX DT 23-JAN-2003 (first entry)
 XX DE Protein #1 related to isomaltose production.
 XX KW Isomaltose.
 XX OS Unidentified.
 XX PN WO200288374-A1.
 XX PD 07-NOV-2002.
 XX PF 25-APR-2002; 2002WO-JP04166.
 XX PR 27-APR-2001; 2001JP-0130922.
 XX PA (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.
 XX PI Kubota M, Nishimoto T, Higashiyama T, Watanabe H, Fukuda S;
 XX PI Miyake T;
 XX DR WPI; 2003-067722/06.
 XX DR N-PSDB; ABV75608.

XX PT Production of isomaltose used for manufacturing e.g. beverages, health
PT foods and cosmetics by using specific non-reducing end
PT alpha-1,4-glucosyl linkage-containing saccharide and alpha-isomaltosyl
PT glucosaccharide synthase -
XX
PS Disclosure: Page 103-111; 121pp; Japanese.
XX
CC The invention relates to a novel method for the production of isomaltose,
CC comprising treating a saccharide having alpha-1,4-glucosyl linkage at the
CC non-reducing end and a glucose polymerization degree of up to 2 with
CC alpha-isomaltosyl glucosaccharide synthase, treating the product(s) with
CC isomaltose-releasing enzyme and collecting the obtained isomaltose.
CC Isomaltose is useful for manufacturing beverages, health foods, feeding
CC materials, cosmetics, drugs and cooking sauces. The sequence represents a
CC polypeptide used in the invention. No further information about the
CC sequence can be obtained from the specification.
XX
SQ Sequence 1284 AA:
Query Match 3.2%; Score 149; DB 24; Length 1284;
Best Local Similarity 18.8%; Pred. No. 0.028;
Matches 170; Conservative 105; Mismatches 321; Indels 308; Gaps 42;
QY 105 TLTYPPSLSPTIYHLDGVEG--GIDFTPHNGTKIINTVAEYNKLSADSGSIHSH- 161
DB 54 TLTVNGAEPDDLIVQVQNGILKVDYRPNISITPSAKTPLMDPNKWTWSAVGATINTTA 113
QY 162 -----LTNNALVEI-----HTANGRWVRDIYLPQGPDLGKRWTFVSAGYSSTVF 207
DB 114 NPMITTSMKIEITKNPVRMTVKKADTTL--FWPESSGGVFDGVRFLHATGDN---M 168
QY 208 YGDKVTLVSGNTLLFKYVNGQWFRSGLENNRITRYAQHINSAELPAHWIVPGLNLVIKQ 267
DB 169 YGIRSF-----NAFDSGGDLLRNSNHAH----- 193
QY 268 GNLGRLLNDIKIGAPGELLHLLIDIGMTTPRDFDAKDEAHREYFQ--TIPVSRMIV 325
DB 194 ---AGEQGD---SGGLTIWSTAGYGLLVSDGYPYTDSTTGQMEFYVGGTPPGRRYA 246
QY 326 NNYAPLHL-----KEVMLPTGELLTDMOPGNGWHSGTNR-----Q 361
DB 247 KONVEYIIMLGPKEIMTDVGE-ITGKPPMLPWSLGFNFNFWDNQETFTNNVDYTRAK 305
QY 362 RTGKELVSHGIDNANYGLNSTAGLGENSHYVVAOLAAHNSRGNYANGIOVHG----- 414
DB 306 NIPIDAYAFDYKKYGETNYGEPAWNTNFPSASTSLAKTMD-AGKIKMIGITKPRIV 364
QY 415 -GSGGGGIYV---LDSTLGNFSEHVGHNHGLHYVDGFGKGSVHRSAE---NNNSTGWGDG 468
DB 365 TKDASANVTTOGTDATNGCYF--YPCHN---EYQDYFIPVTVRSDPYNANERAFWNH 418
QY 469 DK---KREIPNYPYPSOTNEKS-----CLNNCOEPEFDGHKFGFDAMAGSP--- 511
DB 419 STDALNKGIVGNWDETDKSGGALYFNGTNTTGHMSQTHTEGER----ATSGAQRVW 474
QY 512 -----FSAANRF--TMYTP-----NSSAIIQRFENKAVFDSRSSTGFSKWN 551
DB 475 QTARTFYPCQAQYATLWSGDIGIQYNGKGERINWAAGMQ---EQRAVLLSSVNNQGVKWG 531
QY 552 ADT-----QEMEPY-----EHTIDRA 567
DB 532 MDTGGFNODGTNNPNPDILYARMWOFSAITPFRVHGNHNOQROPIYFGSTAEEASKEA 591
QY 568 EQI-----TASVNELSESKMAELMAEY---AVVK--VHMN-CNW----- 601
DB 592 IQRLYSILPYWYAYERSAYENGVLRLQYPTDAAVKNTDAMFCDWLLAAPVVDK 651
QY 602 ---TRNIYIPTASDN--RGSILT-----INHEAGYNSYLFIDGKGVYSGQYKXSF 648
DB 652 QQTSKDIYLPSPGWDIYARGNAITGGQTYRYSVNPDTLMDPLFI-----KKGA 700
QY 649 VSDGQFWKERDVVDTREARKPEQFGVPVTVLGV--YYDPEGTLSYIYPAMYGAYGFTYS 706

DB 701 IIPQ--KVQDYVYGQASVTSVDVDFPDITQSSFTYYDDG-----ASYNKESGTYF 750
QY 707 DDSQNLSDNDCQLQVDTKEQLRFLRANHRANNTVYKFNHNPVTSQPTQATLVCKNKI 766
DB 751 KONMTAODNG-----SGSLFTLGAKSGSYTPALQSYI---VKLHGSAGTSVTNNS- 798
QY 767 LDTKSLTPAPEGLTYTVNGQALPAKENEGCIVSVNSGKRYCLPVGQSGYSLPDMIVGOE 836
DB 799 -----AAMTSYASLALKAAGEG-----NATGKD 823
QY 827 VYVD 830
DB 824 IYGD 827
RESULT 8
ABB47329
ID ABB47329 standard; Protein; 1530 AA.
XX AC ABB47329;
XX DT 05-FEB-2002 (first entry)
XX DE Listeria monocytogenes protein #33.
XX KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
KW vitamin B12; bacterial infection; disease.
XX OS Listeria monocytogenes.
XX PN WO200177335-A2.
XX PD 18-OCT-2001.
XX PF 11-APR-2001; 2001WO-FR011118.
XX PR 11-APR-2000; 2000FR-0004629.
XX PA (INSP) INST PASTEUR.
XX PI Buchrieser C, Frangeul L, Couve E, Rusnok C, Fslhi H, Deboux P;
PI Dusserget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;
PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;
PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
PI Madueno E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;
XX Rose M, Voss H;
XX WPI; 2002-010914/01.
XX PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
PT and prevention of Listeria and related bacterial infections, and
PT related polypeptides -
XX PS Claim 6; SEQ ID No 34; 192pp; French.
XX CC The present invention relates to the genome sequence of Listeria
CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of
CC it are useful for selecting probes and primers for detecting genes in L.
CC monocytogenes and related organisms, and for studying genetic
CC polymorphisms and other genomes. The present sequence is a protein
CC encoded by the genome sequence of the present invention. Proteins
CC expressed from the genome sequence are useful for raising specific
CC antibodies. Identification of L. monocytogenes and related organisms, and
CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin
CC B12. The genome sequence and proteins encoded by it are also useful for
CC selecting compounds that regulate gene expression and cell replication
CC and modulate L. monocytogenes-related diseases. In addition, the genome
CC sequence and proteins encoded by it are useful in pharmaceutical and
CC vaccine compositions for the treatment or prevention of infections by L.
CC monocytogenes and related organisms.

Db	254	SEVDNT-----IAGLSADFP-----VDGAMLDVQWFGGVTADSDDRMG-----TLD	296
Qy	291	DIGMLTTPDRPF-DFAKDKEAHREYFQTIPIVSRMTVNNYAPLHLKEVMLPTGELLTD--	346
Db	297	WDSRFPDPACKIADLAEDGVG-----IPIEESYVGRNLPFHAR--HAADGYLVRSCC	348
Qy	347	MDP-----GNGG-----WHSGTMRQRTIGKELVSHGIDNAN---YGLN	380
Db	349	ATCPVPVYLTGNPWWGKGMIDWTQPEAGAVHDBEQRLVDECVLGHWLDLGEPEMYDPN	408
Qy	381	S-TAGL--GENSH-----PYVVAQLAAHNSRGWYANGI	410
Db	409	DWTAGVIGPKHAHADYHNAYNLLWAQSIADYADNGVQKPEPMLTRAA-----AAGI	460
Qy	411	QVHG--GSGGGGLVTDSTLGNFESHEVGHNY-GLGHY---VDCFKSGVHRSANNSTWG	465
Db	461	QRHGAGMWSADIGSTKRALGSOQAQMHMSGIDYVSGDIGGFREM-ADGDVNLITQ	519
Qy	466	WDGDKRRIPINFPYSQTNKESCLNNCOEPPDGHKFGFDAMAGGSPFSAANRFTMYTPNS	525
Db	520	WPDASAWFDTPLRPHDMLNCLET-----SPDSIGD-----VASNR	556
Qy	526	SAIIORFFENKAVDSRSSTGFSKWNADTOEMEPYEHTI--DRAEQI-----	570
Db	557	ENLVRY-----ELAPYYISLAHRAHQFGEPLAPPLYYYQN	593
Qy	571	TASVNELSESMAELMAEYAVVKVMNGWNTRNIYIPTASADNRGSILTIHHEAGYSNY	630
Db	594	DDHVREMGHQM--LGRDLLIAIVAGEGERRDVYLP-----GEM	632
Qy	631	LFINGDEKVVSGYKKSIVSGQFWKRDVVDREARK-----PSQGVVPVTLVGYVD	684
Db	633	IDITHNERIQSTG--QWIDNVPLW--RDGVFTLPAYARAGAIIPKAFVDASTKDTGKR	687
Qy	685	PEGTLSYYIPAMYG--AYGFT-YSDSQNLSDNCQLQVDTKEGOLRERLANHPANNT	740
Db	688	EDAAVRNELIATVYADVASDFTLYEDOGATTAYADGAVRT---QISQSLTNGVATVT	743
Qy	741	V--MVKFHINVPTESQATLVCNKNKILDTKSLTPAPEGLVTVNGQALPAKEN-----	793
Db	744	VGNASGTYSGAPS-TRPTVVELVDGTQASTVSLGSP--LTEHANKAAFDAASSGWYNA	800
Qy	794	EGCIY-----SVNSGKRYCLPVGQRSGYSL-----PDWIVGQEVYVDSGAKAKVLLSD	841
Db	801	GGGLVVAKAASSSVNTAKTFSFTLIGESVWATFSCENATITFGQYVV-----	848
Qy	842	WDLNSYNRIGEF--VGNVNPADMKVK-----AHNQYLDFSPKSRMRYVK	886
Db	849	-----VGNVPQLGNWSPADAVKLEPSAYPTWTGVVRNL--PPSSTVEWK	890
RESULT 10			
ABG30565			
ID	ABG30565	standard; Protein: 965 AA.	
AC	ABG30565;		
DT	07-OCT-2002	(first entry)	
DE	Alpha-isomaltosylglucosaccharide synthase #3.		
KW	Alpha-isomaltosylglucosaccharide synthase; sugar production;		
KW	cyclic tetrasaccharide; cyclo(-6)-alpha-D-glucopyranosyl(1-3);		
KW	alpha-isomaltosyltransferase; viscosity modifier; sweetener;		
KW	heat stabiliser; hydrolysis stabiliser; colour stabiliser; foodstuff;		
KW	cosmetic; drug composition.		
OS	Unidentified.		
PN	W0200255708-A1.		
PD	18-JUL-2002.		

```

XX 09-JAN-2002: 2002W0-JP00052.
XX PF
XX 12-JAN-2001: 2001JP-0005441.
XX PR
XX (HAYB ) HAYASHIBARA SEIBUTSU KAKAKU.
XX PA
XX Kubota M, Maruta K, Yamamoto T, Fukuda S;
XX PI
XX WPI; 2002-520129/55.
XX DR
XX
XX New alpha-isomaltosylglucosaccharide synthase of bacterial origin for
XX PT the production of cyclic tetrasaccharide gum
XX PS
XX Disclosure; Page 137-141; 144pp; Japanese.
XX
XX The invention describes novel microbial polypeptides having
XX CC alpha-isomaltosylglucosaccharide synthase activity. The proteins are
XX CC useful for producing a sugar (I) having at least three glucose units
XX CC connected by alpha-1,4 bonds except for alpha-1,6 at the non-reducing
XX CC end. The invention also describes a method for the production of the
XX CC cyclic tetrasaccharide cyclo(-6)-alpha-D-glucopyranosyl(1-3)-alpha-D-
XX CC glucopyranosyl(1-6)-alpha-D-glucopyranosyl(1-3)-alpha-D-glucopyranosyl
XX CC (I-) (II) by treatment of (I) with alpha-isomaltosyltransferase.
XX CC (II) and similar sugars in crystalline or syrup form are used as
XX CC a viscosity modifier, sweetener, heat stabiliser, hydrolysis stabiliser
XX CC and colour stabiliser for foodstuffs, cosmetics and drug compositions.
XX CC This is the amino acid sequence of an alpha-isomaltosylglucosaccharide
XX CC synthase.
XX
XX Sequence 965 AA;
XX SQ
Query Match 3.1%; Score 145; DB 23; Length 965;
Best Local Similarity 19.8%; Pred. No. 0.037;
Matches 165; Conservative 107; Mismatches 288; Indels 272; Gaps 49;
QY 176 RWRDYILPGQ---PDLEGKMRVFSAGYSSTVFYGDGRKVTLSVGNTLLFKFYNGQWFR 232
DB 246 RW-----YLMGDDLPDLRHDMELTGTPTPPPKAPG-----LWSE---FGYDN--W-- 289
QY 233 SGELENNRITYAQHIGWAEPLPAHWIPLGLNLVIKQ-GNLGRLNDIKIGAPGELLHTI- 290
DB 290 -SEVDNT-----IAGLRADFP---VDGMLDVQWFGVTADSDTTRMG-----TLD 332
QY 291 -DIGMLTTPDRF--DFAKDKEAHREYFQTIPVSRMIVNNTAPLHLKEVMLPTGELLTD-- 346
DB 333 WDSRFPDPACKIADLAEDVG-----IPIESYVGRNLPEHAR--MAADGYLVRS GC 384
QY 347 -----MDP--GNGG-----WHSCTMRQRIKGVLSHGIDNAN---YGLN 380
DB 385 ATCPVPVYLTGNPMWKGGMIDWTQPEAGAVNHDQRQLHDEGVGLHWDLGEPEMTDPN 444
QY 381 S-TAGL--GENSH-----PYYVAQLAHNSRCNYANGI 410
DB 445 DWTAGTVPGRHAAHYINAYNLLWAQSIADGYADNGVKRPFMLTRAA-----AAGI 496
QY 411 QVHG--GSGGGIVTLDSTLGNEFSHEVGHNY-GLGHY---VDGFKGSVHRS AENNSTWG 465
DB 497 QRHAGWSADIGSTMKALGSQQAQMHMSMSGIDYIGSDIGGFREM--ADGDVNELYTQ 555
QY 466 WDGDKKRFIPNTFVPSQTNEKSLNNQCQEPFDGKHFGFDAMAGGSPFSAANRFTMYTPNS 525
DB 556 WFADSADFDTPLRPHTDNLNCLET-----SPDSIGD-----VASNR 592
QY 526 SAIIOREFENKAVFDSRSSTGFSKNADTQMEPEYHTI-DRARQI----- 570
DB 593 ENLVRRY-----ELAPYYISLAHRAHQFGEPLAPLVITYQN 629
QY 571 TASVNELSESKMAELMAEYAVPVHMMNGNWNTRIYIPTASADNRGSGILTINHEAGNSY 630
DB 630 DDHVREMGHQK---IGRDLLIIAIVAGEGEREDVYILPA-----GEW 668
QY 631 LFIGNDEKVVYSGYKKFSVDGQFWKREDVYDTREARK-----PEQGVPTVLVGYD 684

```

Db 669 IDIHTNERIQSTG---QWIDNVPWM--RDGVFTLPAYARAGALIIPRAVDASTKIDITGKR 723
 Qy 685 PEGTSLSYIYPAMYG---AYGFT-YSDSQNTSDNCQQLQVDTKEQLRFLRANHRANT 740
 Db 724 EDAVRNELIATYADVADSDFTLYEDDAGATTAYADGAVRTT---QISQSLTNGVATVT 779
 Qy 741 V--MNKFHINVTESOPTOATLVCCNKKILDTSKLSLTPAEGTLTVNGQALPAKEN----- 793
 Db 780 VGRASGTYSAPS-TRPTVVELYDTGTQASTVSLGVP--LTHANKAAFDASSGWYNA 836
 Qy 794 -EGCIV-----SVNSGKRYCLPVGQRSGYSL-----PMIVGQEVYVDSGAKAKVLLSD 841
 Db 837 GGGLVAKAASSSVNTAKTFSFTLGESVWATFSCENATTTFCQSYV----- 884
 Qy 842 WDNLNRYRGEF--VGNVNPADMKKYK-----RWNGQYLDKSPSRMRVYK 886
 Db 885 -----VGNVPOLGNMSPADAVKLEPSAYPTWTGVVRNL--PPSSTVEWK 926
 RESULT 11
 AAB10667
 ID AAB10667 standard; Protein: 2057 AA.
 AC AAB10667;
 XX
 XX 19-JAN-2001 (first entry)
 XX
 XX L. mesenteroides alternan sucrose protein.
 DE
 XX Alternan sucrose; glucosyltransferase; fructose; cosmetic; foodstuff;
 KW syrup.
 KW
 XX Leuconostoc mesenteroides.
 OS
 XX DE19905069-A1.
 PN
 XX 10-AUG-2000.
 PD
 XX 08-FEB-1999; 99DE-1005069.
 PF
 XX 08-FEB-1999; 99DE-1005069.
 PR
 XX (PLAN-) PLANTEC BIOTECHNOLOGIE GMBH.
 PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
 XX
 XX Kossmann J, Welsh T, Quanz M, Knuth K;
 PI
 XX WPI: 2000-550294/51.
 DR N-PSDB; AAA97904.
 DR
 XX New nucleic acid encoding recombinant Leuconostoc mesenteroides
 PT alternan sucrose protein and methods of alternan and fructose
 PT production -
 XX
 XX Claim la; Page 30-36; 64pp; German.
 PS
 XX This invention describes a novel nucleic acid molecule (I) encoding an
 CC alternan sucrose (E.C. 2.4.1.140 - an enzyme, that belongs to the
 CC glucosyltransferase group) The recombinant, purified alternan sucrose
 CC gene is useful for the fermentative production of alternan (a
 CC carbohydrate) and/or fructose by secreting the enzyme into a
 CC saccharose-containing culture medium. Alternatively, the enzyme is
 CC contacted with a saccharose-containing solution. The alternan and/or
 CC fructose is then isolated from the medium. Cosmetic products or
 CC foodstuffs containing alternan can be produced. Recombinant production of
 CC alternan sucrose is advantageous as it provides a cost effective means of
 CC producing fructose for high fructose containing syrups, production of
 CC which previously has been achieved by costly production from maize
 CC starch. This sequence represents the Leuconostoc mesenteroides alternan
 CC sucrose protein which is described in the method of the invention.
 XX
 XX Sequence 2057 AA;
 SQ

Query Match 3.1%; Score 144.5; DB 21; Length 2057;
 Best Local Similarity 17.6%; Pred. No. 0.14; 346; Indels 485; Gaps 52;
 Matches 203; Conservative 122; Mismatches 122; Indels 485; Gaps 52;
 Qy 22 TAADNNSAIYFNTSQPINDLOGSLAAEYKFAQSILPAHPKEGDSQPHLTSLRKSLL-- 79
 Db 910 TTTQDNSQDYKN--QGIGVIVGN-NPDLKLNNDKITLHM-----GKAHKNOLYRALVLSN 962
 Qy 80 ---VRPVXADKTPVQVEARDNNKIL-----GTLTLYP-PSLPTPTIHLDG-- 123
 Db 963 DSGIDVYSDDKAPT-LRTNONGDLIFHKTNTFKQDCTIINYEMKGLNALISGYLGW 1021
 Qy 124 VPEGGIDTPHNGTKKIINTVAENVKLSDAGSSISHS--LTNNALV----- 169
 Db 1022 VPVGASD-----SQDARTVATESSSN-DGSVTHSNAALDSNVIEGFSNQAMTSP 1073
 Qy 170 -----INTANGRWYRDI-----YLPQGPDLCKMVRVSS-----AGYSSTVFYGD 210
 Db 1074 EQSTNVVIATKANLKFELGITSFELAPOYRSSGDTNYGCM-SFLDSFLNGYAFTRDY-D 1131
 Qy 211 RKVTLVSGNTLLFKYVNGQWFRSGELENNRITYAQHIWSAELPAHWI-----VPLGLNV 264
 Db 1132 LGFNKADGNPNETKYGTDDLRNA-----JEAUHKNGMAIADWVPDQIYALPGKEVV 1184
 Qy 265 I-----KQGN-----LSG-----RLNDIKI 279
 Db 1185 TATRVDERGNQLKDTDFVNLVYVANTKSSGVYQAKYGGEFLDKLREEYPSLFKQNOYST 1244
 Qy 280 GAP-----GELLHTIDIGMLTTPDRPDRFAKOEAREHREYFOTIPVSRM 323
 Db 1245 GQPIDASTKIKOWSAKYMNGTNIH-----RGAYYYVKD-WATNQYFNIAKTREV 1293
 Qy 324 IVNNTAPLHLKXVMLPTGELL-----TMDPGNGGWHSGTMRQRICKE 366
 Db 1294 ---FELIQLQNKDAQTGFISDASGVKYYSISGYQAKDTFFIEDGNGNY-- 1339
 Qy 367 LVSHGIDNANYGLNSTAGIGENSHPYVVAQLAAHNSRGN-----ANGIOVHGSGGGGIVT 423
 Db 1340 ---FDKGYMVRSQG---ENPIRTVETSVNTRNGYFMPNGVELRKFG----- 1384
 Qy 424 LDSTLGNESHEVGHNYGLGHYVDGFKGSVHSAENNSNTHGWDGDKKRFIPNFYPSQTN 483
 Db 1385 -----TNSGNVYTFDDOGKRVDRKYINDDAN 1411
 Qy 484 EKSCNLNOCQEPFDG-----HKFGFDAMAGSPFSAANRFTM---YTPNSSAIIQREFE 534
 Db 1412 NFYHLN-----VDGTMGRGLFKFSDTL---QYFASNGVQIKDSYAKDSKG----- 1454
 Qy 535 NKAVFDS---RSSTGFCKWADTQEMEPYEHYIDRAEQITASVN----- 575
 Db 1455 NKYFDSATGNNDTG---KAQTDWNGYIITSDANNITGVNTDYATYITSSLEDGL 1510
 Qy 576 -----ELSEKMAELMAEYAVVYVVMWNGNTRNIYIPT 609
 Db 1511 PANAPYGVVTKDQNGNDLAKQYINHTKYEQGOOVQVTRQYTDTSKGVSWN-----LITF 1563
 Qy 610 ASADNRGSILTINHE-----AGNYSYLFINGDEKYVSGYKKSFYSD 651
 Db 1564 AGDLOGQLRWDSRALTWTPTKTNQOISFIYANRNDGLFLNAPYQV--KGYQLAGSN 1621
 Qy 652 -----GQFWERDVVDTRARKPEQ----- 671
 Db 1622 QYKGOQVTIAGVANVSGKDWLSISFNGTQYITDSQALNTNTHDMNQVFTVNTTSLNDGL 1681
 Qy 672 -FGVP-----VTLVGYVDPSTLSSYIYPAMYGAYGFTYSDSQN 711
 Db 1682 FLNAPTRQGYKLAGLAKYNNQYTVSQYFDDGTVGVVGLGQVTVWDNHAQMO 1741
 Qy 712 LSDNDCQLQVDTKEGOLRFLANHRANNTVMKHFHINVPESQPTQATLVCNKKILTDS 771
 Db 1742 VSDTDOQLVYN-----SNGRNDGLFLNAPYRQGSQLI----- 1774

24-APR-2001 (first entry)
M. catarrhalis les1 200kDa protein SEQ ID NO:11.
Moraxella catarrhalis strain Q8; major outer membrane protein;
200kDa outer membrane protein; antibacterial; immunogenic; infection;
otitis media; detection.
Moraxella catarrhalis.
WO200107619-A1.
01-FEB-2001.
26-JUL-2000; 2000WO-CA00870.
27-JUL-1999; 99US-0361619.
(CONN-) CONNAUGHT LAB LTD.
Loosmore SM, Sasaki K, Yang Y, Klein MH;
WPI; 2001-159722/16.
N-PSDB; AAF59105.
New nucleic acid encoding Moraxella catarrhalis outer membrane protein,
useful in protective vaccines and for diagnosis -
Claim 1; Fig 5A-Y; 247pp; English.
The present invention describes an isolated and purified nucleic acid (I)
that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.
The 200 kDa outer membrane protein (II) has antibacterial activity and
can be used in vaccines. (II), and its truncated versions, are used as
immunogenic compositions and vaccines to protect against M. catarrhalis
infections, particularly otitis media in humans. (II) is also used as
antigen in immunoassays for detecting specific antibodies (Ab), and to
generate Ab. (I) are used for recombinant production of (II) and its
fragments are used as probes for identifying/cloning 200 kDa protein
genes from other strains, and for diagnostic detection of M. catarrhalis.
(I) makes possible production of large amount of recombinant immunogens.
Expression of truncated versions of (II) reduces toxicity of the protein
towards the Escherichia coli host. The present sequence represents the
M. catarrhalis les1 200kDa protein, which is given in the exemplification
of the present invention.

Query Match 3.1%; Score 144; DB 22; Length 2314;
Best Local Similarity 20.1%; Pred. No. 0.18;
Matches 212; Conservative 129; Mismatches 388; Indels 326; Gaps 52;

QY 9 TILAPLAIGVFSATA-----ADNNSAIVNTSOPINDLQGSIAAEVKEFAQOILPAHPKE 63
DB 47 TRVATLAILVIGATLINGSAYAQNNSKIAFGTGG--NNDNASASNEASIAIGSLAKARANQ 104
QY 64 ----GDSOP-----HLTSLRKSLLVPRVADDKTPVQVEARD---DNKKIL 103
DB 105 ATAIGSRPDRPQAQNAQKAGSHAGK-RESIAIGSDVLAEGDASIAIGSDDLIDRN--- 160
QY 104 GRLTYPPSSLPTTYHLIDGVEGGIDPTPHNGTKKIINTVAENKLSADSGSSIHSS--H 161
DB 161 STNSKYPNGULLSTLQNTHTVLQ-----IPDSNGSOKYRRTAAE-GHASTAYGAMAYAKGH 215
QY 162 LTNALVEIHTANGWRVDIYLPQGDLEGKMYRVSSA---GYSTFYVGDVKVTLVSG 218
DB 216 FA-NAPGTRSTAEAGNYSIAVGUTAKAE-KGYTIAIGSNAQAINYGALGADTRVDLDYG 273
QY 219 NTILF-----KYV---NGWFRSGELENRRITYAQHIWISAEPLPAHWIPLG 261
DB 274 IALGYSQILNNNNNNNNKAYPEGNGSNIKSKATGNCLF---SIGSSTIKRKLIINVA 330
QY 262 -----NLVFIKGNLSGRINDI-----KIG-----APGELLHTID 291

DB 331 GYEDTDVAVNAQLKAVENLAKRQITFKGDDNGTVGKKKLGETLTIKGETQADKLNNN 390
QY 292 IGLWLTTPRD--RFDFAKDKEAREYFQTPVSRMIV-----NNYAPLHLREY-MLP 339
DB 391 IGVVTDNNTGLKVLAKNLG-----LETVSTKNLTASEKVTGSGNNTAELOSGLTFTP 446
QY 340 TCELLTD-----MDPONGGWBHSGTHRQRIGKELYSHGIDNANYGLNSTAGLGEN 388
DB 447 TTNASTDKTVYGTGDKLFTDNTALEDTT---RITDKI-----GFSNKAQTVDE 494
QY 389 SHPYV-VAQLAHNSRGNVANGIQVHGSGGIGVITLDSTLGNFESHEVGHNYGLGHVVD 447
DB 495 NKPYLDKDKLVGNSTLN-----NGG---LTVNNTIG-----GSKKOIQVAD 534
QY 448 GFKGSVHRSANNNSTWGN-----DG-----DKRF----- 473
DB 535 GIKFADVNVNVSNAKFGTTRITEBEIGFADAGKVKRKSPLYLDKKQLQGVGVKITKDSG 594
QY 474 -----IPNFYPSOTNEKSLN---NOCOPPDGHKFGFDAM-AGGSPFSAANRFTM- 520
DB 595 INAGDOKISN-VKDATDDDDAVTYKQLKQVQDADGALQSFSIRDEKQGETISNLSNG 653
QY 521 YTPNSAIIQREFENKAVFDSRSSTGFSKWNADTOEMEPYEHTIDRABQITASVNELS 580
DB 654 NTPNTFETITFAGENGISISNDIAKGVKVID-----PINGLTTP 694
QY 581 KM---AELMAEYAVVVMWNGNWTFRNI---YIPT-ASADNRGSILT----- 620
DB 695 KLTVGSDKDKTQLVIEQVAGSNDTKNIIRGLSPTLPSITNAGGVRTTEQGNWITSDEK 754
QY 621 -----INHEAGVNSYLFINGDEKVVVSQGY-----KKSFVSDQGF 654
DB 755 SKAASIGDILNTGFNLKNNNSVGFVSITYNTVDYFDGNATTAKVTYDETNTSKVTYDVN 814
QY 655 WBERVVDVTRAKRPQGPQVPTTLVGYDPPGTLSSYIYPAMYGAYFTYSDSDSLS 714
DB 815 VDEKTIETLGDNGKTKNIGVKTTLT-----TTNA-----NGKATNFST 853
QY 715 NDCQLQVDTKEGRLPRLANHRANNTVMKHNIVPTESQPTQATLVCMNKILDTKSLTP 774
DB 854 TDNDALVNAKD-----IAENL--NTLAKETHTGTADTALQTFVKKDKGATDEITIV 905
QY 775 APBGLTYTVNGQ---ALPAKRENGCIVSYNSGKRYCLPVGQRSGYSLPDMIVGQEVYVDS 831
DB 906 GKDG---TQNGKTVNTLKLKGENGLVATNK-----DGTVTGINTQS 945
QY 832 GAKAKVLLSDWMLSYNRIGEFVGNVNPADMKVK 866
DB 946 GLKA-----GDSITLTKDG--LSIKNPASNEQIQ 972
RESULT 14
ABG30538
ID ABG30538 standard; Protein; 1251 AA.
XX
AC ABG30538;
XX
DT 07-OCT-2002 (first entry)
XX
DE Alpha-isomaltosylglucosaccharide synthase #2 mature protein.
XX
KW Alpha-isomaltosylglucosaccharide synthase; sugar production;
KW cyclic tetrasaccharide; cyclo(-6)-alpha-D-glucopyranosyl(1-3);
KW alpha-isomaltosyltransferase; viscosity modifier; sweetener;
KW heat stabiliser; hydrolysis stabiliser; colour stabiliser; foodstuff;
KW cosmetic; drug composition; enzyme.
XX
OS Unidentified.
XX
PN WO200255708-A1.
XX
PD 18-JUL-2002.

[illegible]

CC probiotic specimens, probiotic-containing food, feed and fodder, by
CC introducing the plasmidic gene into the structure of the bacterial
CC strain, preferably L.lactis IBB500, IBB501, IBB502, IBB140 and
CC subsequently using the obtained bacterial strain in plant material and
CC milk fermentations or in biomass production. The present sequence
CC represents the lactococcus amylolytic enzyme.
XX
SQ Sequence 784 AA;
Query Match 3.0%; Score 142; DB 24; Length 784;
Best Local Similarity 20.0%; Pred. No. 0.046;
Matches 141; Conservative 95; Mismatches 283; Indels 186; Gaps 36;
QY 239 NRITTAQHIAELPAHIVPGLNLVKGSLGRLNDIKIGAPCELLLHTIDIGMLTIP 298
DB 122 NHATNTIGWTLTVPGN--QNGHYAYKLTADGTVDSDYAGSTYGTISTSVS-----NTT 175
QY 299 RDRFDPAKDKEAHREYFOTIPVSRMIVNNYAPLHLKEVMLPTGELLTMDPGNGCWHSGT 358
DB 176 NDPYSIATTQGNRSVVE-----SANLASNLA-----LAQKKSAT-----WRVAS 216
QY 359 MRORIGKELVSHGIDNANYGLNSTAGLGENSHPYVQAQLAHNSRGNYANGIQVHGSGG 418
DB 217 PTOALVDEL--HIRD---FTSSSTSGV-----SAGNRKFLGVIOQSGTTDPN 258
QY 419 GGIVTLDSTLGNESHEVGHNYGLCHYVD---GFKGSVHRSANNNSTWGDGDKRRFI 474
DB 259 TGTAT-----GLDYLKNEGFNYIOLHASQSVGVGSGNRRTTAQPNFNMGYDPQMKWYQ 313
QY 475 PNFYPSQTNKESCLNOCQEPFDG-HKFGFDAMAGGSPFSAANRFTWYTPNSSAIIORFF 533
DB 314 SEYASNSVNPVTRI-NEMKEMVQGLHTNGISVYMD---MYLNH--VYOSASAFEKA-- 364
QY 534 ENKAVFDSRSSTG-----FSKNADTQEMEPYEHTID--RAEQITASVNELSES 580
DB 365 EPGYVFRKNTQSGCGNDTASNHEMFGKYIIDSVTYWAKNYDIDGFRDEMTL----LDST 420
QY 581 KMAELMAEYAVVKVHM-----WNGNWTFRNIYIPTASADNRGSILTINHEAGYNSYLFIN 634
DB 421 TNKLRAALTALDPHIIIMYGEGWGDSDNANN--IPETSINNYKNVPGIG-----FFNP 470
QY 635 GDEKVVSO--GYKKSFSVSGQFWMKRDVVDTREARKPEQFQVPVTVL-----GYVD 684
DB 471 GERDAISNNGSAGGFAA-----GNTASTITVACALLASGGWN 508
QY 685 PEGTLLSSYIYPAMYGAY-----GFTYSDSQNLSDNCQLQVDTREGQLRFLANHRANN 739
DB 509 GNGTVQAFLTQSINMYVECHDSFTLNDLSLADPND-----SVATHQARY 554
QY 740 TVYMKFHI--NVPTEOQTATLVCNNKILDTKSLTAPAGILT--YTVNGQALPA----- 790
DB 555 TLNATNINLNGVTPMETGQE--FDQSKLVNPSNLTPLSPTQAYQSGSMEKPAWYPAS 612
QY 791 -----KENEGCIVSVNGKRYCLPVGORSYSLPDWIVGOEYVVDGSAKAKVLLSDMDNL 845
DB 613 WDTAKNSYNGLFLASNCTYY-----GNTWP---GSNLYTPVYAGDVVNAMWMDNV 660
QY 846 SYNRIIG-EFVGNVNPADMKKVAHWNGYL--DFSK-----PRSMRV 883
DB 661 KDNQNAVNFIGN-----LMKFKKSNPQFWDYDKSLAWTPTSIGV 700

Search completed: October 8, 2003, 16:24:44
Job time : 59.9419 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 8, 2003, 16:19:57 ; Search time 47.5173 Seconds
(without alignments)
4811.606 Million cell updates/sec

Title: US-10-002-309b-2

Perfect score: 4709

Sequence: 1 MKLYLSTLPLAIGVFS.....ANNGQYLDSPSRMRVYK 886

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_23:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phage:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_rvirus:*

16: sp_bacteriap:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4709	100.0	898	082882	082882 escherichia
2	1548	32.9	310	092AL1	Q92all escherichia
3	686.5	14.6	1002	09AGX6	Q9agx6 vibrio chol
4	671	14.2	1335	16 Q9KN18	Q9kn18 vibrio chol
5	571	12.1	875	16 Q8D5P4	Q8d5p4 vibrio vuln
6	290	6.2	957	16 Q9KTH2	Q9kth2 vibrio chol
7	166.5	3.5	1217	16 Q97GW3	Q97gw3 clostridium
8	161	3.4	1038	2 Q45780	Q45780 bacteroides
9	159.5	3.4	2367	2 Q46034	Q46034 clostridium
10	158.5	3.4	2204	5 Q81L70	Q81l70 plasmodium
11	156.5	3.3	1157	16 Q97F62	Q97f62 clostridium
12	154.5	3.3	2367	2 Q9F931	Q9f931 clostridium
13	150	3.2	3692	2 Q8KRR3	Q8krr3 fusobacteri
14	149	3.2	1284	2 Q8RQU9	Q8rqu9 bacillus gl
15	148.5	3.2	2468	16 Q912M3	Q912m3 pseudomonas
16	148	3.1	1021	2 Q46085	Q46085 clostridium

17	148	3.1	2937	16 Q8D990	Q8d990 vibrio vuln
18	147.5	3.1	1530	16 Q8Y479	Q8y479 listeria mo
19	147.5	3.1	2306	16 Q8F107	Q8f107 leptospira
20	147	3.1	1093	16 Q8EV69	Q8ev69 mycoplasma
21	147	3.1	1377	16 Q98Q28	Q98qz8 mycoplasma
22	147	3.1	1386	16 Q92DL0	Q92dl0 listeria in
23	146.5	3.1	1196	17 Q8TS48	Q8ts48 methanosarc
24	146.5	3.1	2057	2 Q9RE05	Q9re05 leuconostoc
25	145.5	3.1	2189	5 Q8IRK6	Q8irk6 plasmodium
26	143.5	3.0	1395	2 Q9AISO	Q9ais0 staphylococ
27	143	3.0	1284	16 Q8EV71	Q8ev71 mycoplasma
28	140	3.0	1135	1 Q9P9D1	Q9p9d1 uncultured
29	139	3.0	1110	17 Q8PXT0	Q8pxt0 methanosarc
30	138.5	2.9	1270	16 Q8XAN9	Q8xan9 escherichia
31	138.5	2.9	1575	2 Q9LCH3	Q9lch3 streptococ
32	138	2.9	958	17 Q8TL14	Q8tl14 methanosarc
33	138	2.9	1964	2 Q8KQK9	Q8kqm9 moraxella c
34	138	2.9	2675	5 Q813F9	Q813f9 plasmodium
35	138	2.9	2894	17 Q58791	Q58791 methanococ
36	137.5	2.9	2529	16 Q25579	Q25579 helicobacte
37	137	2.9	1480	2 Q8GC13	Q8gc13 lactobacill
38	137	2.9	5559	16 Q82KG6	Q82kg6 salmonella
39	136.5	2.9	1457	2 Q9F289	Q9f289 versinia pe
40	136.5	2.9	3705	2 Q9F285	Q9f285 versinia pe
41	136.5	2.9	3705	16 Q8ZHA1	Q8zhal versinia pe
42	136.5	2.9	3710	16 Q8CZU2	Q8czu2 versinia pe
43	136	2.9	1677	16 Q8RGN7	Q8rgn7 fusobacteri
44	136	2.9	1980	3 Q05747	Q05747 saccharomyc
45	136	2.9	2167	16 Q92EK5	Q92ek5 listeria in

ALIGNMENTS

RESULT :

082882 ID 082882 PRELIMINARY; PRT: 898 AA.
AC 082882: Q9ZGUL;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Plasmid POL157 DNA, complete sequence.
GN L7031.
OS Escherichia coli O157:H7.
OG Plasmid POL157.
OC Bacteria, Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7;
RX MEDLINE=98290540; PubMed=9628576;
RA Makino K., Ishii K., Yasunaga T., Hattori M., Yokoyama K.,
RA Yutsudo H.C., Kubota Y., Yamauchi Y., Iida T., Yamamoto K., Honda T.,
RA Han C.G., Ohtsubo E., Kasamatsu M., Hayashi T., Kuhara S.,
RA Shinagawa H.;
RT "Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an
RT enterohemorrhagic Escherichia coli O157:H7 derived from Sakai
RT outbreak";
RL DNA Res. 5:1-9(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=EDL933;
RX MEDLINE=98391744; PubMed=9722640;
RA Burland V., Shao Y., Perna N.T., Plunkett G., Sofia H.J.,
RA Blattner F.R.;
RT "The complete DNA sequence and analysis of the large virulence plasmid
RT of Escherichia coli O157:H7";
RL Nucleic Acids Res. 26:4196-4204(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=EDL933;
RA Brunder W.;

RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB011549; BAA31157.3; -;
DR EMBL; AF074613; AAC70099.1; -;
DR EMBL; Y11831; CAA72517.1; -;
KW Hypothetical protein; plasmid.
SQ SEQUENCE 898 AA; 99548 MW; 3C1AE23E3EAE1FAB CRC64;

Query Match 100.0%; Score 4709; DB 2; Length 898;
Best Local Similarity 100.0%; Pred. No. 1.4e-300;
Matches 886; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLYLSTILAPLAIGVFSATAADNNSAIYFNTSOPINDLOGSLAAEVKFAQSILPAH 60
DB 13 MKLYLSTILAPLAIGVFSATAADNNSAIYFNTSOPINDLOGSLAAEVKFAQSILPAH 72
QY 61 PKEGDSOPHLTSLRKSLLLVPRVADKTPVQVEARDNNKILGTITLYPPSSLPDTIYH 120
DB 73 PKEGDSOPHLTSLRKSLLLVPRVADKTPVQVEARDNNKILGTITLYPPSSLPDTIYH 132
QY 121 LDGVPPEGIDFTPHNGTKKIINTVAEVNKLSDAGSSIHSHLTNNALVEIHTANGRWVRD 180
DB 133 LDGVPPEGIDFTPHNGTKKIINTVAEVNKLSDAGSSIHSHLTNNALVEIHTANGRWVRD 192
QY 181 IYLPQGPDLGKMRVFSAGSYSTFYGDKRVTLSVGNLTLLFKYVNGQWFRSGELENR 240
DB 193 IYLPQGPDLGKMRVFSAGSYSTFYGDKRVTLSVGNLTLLFKYVNGQWFRSGELENR 252
QY 241 IYLPQGPDLGKMRVFSAGSYSTFYGDKRVTLSVGNLTLLFKYVNGQWFRSGELENR 300
DB 253 IYLPQGPDLGKMRVFSAGSYSTFYGDKRVTLSVGNLTLLFKYVNGQWFRSGELENR 312
QY 301 RFDAKDEAHREYFOTIPVSRMIVNMYAPLHLKEVMLPTGELLTDMDPGNGGHSCTMR 360
DB 313 RFDAKDEAHREYFOTIPVSRMIVNMYAPLHLKEVMLPTGELLTDMDPGNGGHSCTMR 372
QY 361 QRIGELVSHGIDNANYGLNSTAGLGENSHYVYVAQLAAHNSRNYANGIQVHGSGGGG 420
DB 373 QRIGELVSHGIDNANYGLNSTAGLGENSHYVYVAQLAAHNSRNYANGIQVHGSGGGG 432
QY 421 IYVLDSTLGNFSEHVGHNGLGHVYDQFGKSVHRSANNNSTWGDGKKRFPNFPYS 480
DB 433 IYVLDSTLGNFSEHVGHNGLGHVYDQFGKSVHRSANNNSTWGDGKKRFPNFPYS 492
QY 481 QNNEKSLNNOCEPFDGKHGFGFDMAGGSPFSAANRFTMTTPNSSAIIOFFENKAVFD 540
DB 493 QNNEKSLNNOCEPFDGKHGFGFDMAGGSPFSAANRFTMTTPNSSAIIOFFENKAVFD 552
QY 541 SRSGTGFKNWADTOEMEPYEHTIDRAEQITASVNELSEKMAELMAEYAVVYVHMNGN 600
DB 553 SRSGTGFKNWADTOEMEPYEHTIDRAEQITASVNELSEKMAELMAEYAVVYVHMNGN 612
QY 601 WTRNIYIPTASADNRGSLITINHEAGYNSYLFINGDEKVVSGQYKSPVSDGQFWKERDV 660
DB 613 WTRNIYIPTASADNRGSLITINHEAGYNSYLFINGDEKVVSGQYKSPVSDGQFWKERDV 672
QY 661 VDTREARPEQPGVPTTLVGVYDPEGTLSSYIYPAMYGAYFTYSDSDNSDNDCOLQ 720
DB 673 VDTREARPEQPGVPTTLVGVYDPEGTLSSYIYPAMYGAYFTYSDSDNSDNDCOLQ 732
QY 721 VDTKEQLFRLANHRANTVYVNHKFIHNPVTEQPTQATLVNCKNLTSLTAPAGLT 780
DB 733 VDTKEQLFRLANHRANTVYVNHKFIHNPVTEQPTQATLVNCKNLTSLTAPAGLT 792
QY 781 YTVNGQALPAKENEGCIVSVNSGKRYCLPVGQSGYSLPDMIVGQYVYDVGAKAKVLLS 840
DB 793 YTVNGQALPAKENEGCIVSVNSGKRYCLPVGQSGYSLPDMIVGQYVYDVGAKAKVLLS 852
QY 841 DWDNLNRYGIFGVGNVPADMKKVKANNGQYLDPSKPSRMYVYK 886
DB 853 DWDNLNRYGIFGVGNVPADMKKVKANNGQYLDPSKPSRMYVYK 898

RESULT 2

Q92AL1
ID Q92AL1 PRELIMINARY; PRT: 310 AA.
AC Q92AL1
DT 01-MAY-1999 (TREMELrel. 10, Created)
DT 01-MAY-1999 (TREMELrel. 10, Last sequence update)
DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)
DE Hypothetical 34.0 kba protein (Fragment).
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EDJ933;
RA Brander W.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y11275; CAA72142.1; -;
KW Hypothetical protein; Plasmid.
FT NON_TER 310 310
SQ SEQUENCE 310 AA; 34032 MW; 2AA72C166F315BA6 CRC64;

Query Match 32.9%; Score 1548; DB 2; Length 310;
Best Local Similarity 100.0%; Pred. No. 1.2e-93;
Matches 298; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKLYLSTILAPLAIGVFSATAADNNSAIYFNTSOPINDLOGSLAAEVKFAQSILPAH 60
DB 13 MKLYLSTILAPLAIGVFSATAADNNSAIYFNTSOPINDLOGSLAAEVKFAQSILPAH 72
QY 61 PKEGDSOPHLTSLRKSLLLVPRVADKTPVQVEARDNNKILGTITLYPPSSLPDTIYH 120
DB 73 PKEGDSOPHLTSLRKSLLLVPRVADKTPVQVEARDNNKILGTITLYPPSSLPDTIYH 132
QY 121 LDGVPPEGIDFTPHNGTKKIINTVAEVNKLSDAGSSIHSHLTNNALVEIHTANGRWVRD 180
DB 133 LDGVPPEGIDFTPHNGTKKIINTVAEVNKLSDAGSSIHSHLTNNALVEIHTANGRWVRD 192
QY 181 IYLPQGPDLGKMRVFSAGSYSTFYGDKRVTLSVGNLTLLFKYVNGQWFRSGELENR 240
DB 193 IYLPQGPDLGKMRVFSAGSYSTFYGDKRVTLSVGNLTLLFKYVNGQWFRSGELENR 252
QY 241 IYLPQGPDLGKMRVFSAGSYSTFYGDKRVTLSVGNLTLLFKYVNGQWFRSGELENR 298
DB 253 IYLPQGPDLGKMRVFSAGSYSTFYGDKRVTLSVGNLTLLFKYVNGQWFRSGELENR 310

RESULT 3

Q9AGX6
ID Q9AGX6 PRELIMINARY; PRT: 1002 AA.
AC Q9AGX6
DT 01-JUN-2001 (TREMELrel. 17, Created)
DT 01-JUN-2001 (TREMELrel. 17, Last sequence update)
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
DE ToxR-activated gene A protein.
GN TAGA
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=J95;
RA MEDLINE=21117025; PubMed=11179381;
RA Karaolis D.K.R., Lan R., Keper J.B., Reeves P.R.;
RT *Comparison of Vibrio cholerae pathogenicity islands in Sixth and
RT Seventh Pandemic Strains.
RL Infect. Immun. 69:1947-1952(2001).
CC -1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
DR EMBL; AF325733; AAK20748.1; -;
DR Interpro; IPR003961; FN_III.
DR Pfam; PF00041; fn3; 1.
DR SMART; SM00060; FN3; 1.

KW Hydrolase. 1002 AA; 114645 MW; 6A239DFB6408407F CRC64;
SQ SEQUENCE 1002 AA; 114645 MW; 6A239DFB6408407F CRC64;
Query Match 14.6%; Score 686.5; DB 2; Length 1002;
Best Local Similarity 24.0%; Pred. No. 2.8e-36;
Matches 230; Conservative 157; Mismatches 275; Indels 295; Gaps 40;
QY 27 NSAIYFNTSOPINDLOGSLAAEVKFAQSOILPAHPKEDSQPHLTSLAKSLLLVLP-VKA 85
DB 137 NEIDYVHTTEIKNYSLEGEVETVQTHVI--SPEGRKNEPEITGRDALILFKPSIKN 194
QY 86 DDKTPQVQEARDDNKILGTFLYPPSSLPDTIYHLDGVPEGGIDFTPHNCTKKLIINTVA 145
DB 195 SSSILMKIYSEOG---LTSKVYMKSPSMLPKTDQPID-----ID----- 230
QY 146 EYVKLSASGSSIHSHLTNNALVEIHTANGRWVRDIYLPQGDLEGKMYRFVSAGYSST 205
DB 231 ----- 230
QY 206 VFYGDVKVTLVGNITLLFYVNGWFRSGELENRR-ITYAQHINSAELPAHWIVPGLNIV 264
DB 231 -----ENNVKVSYSNSYSAEIPWKKMKGMSLH 259
QY 265 IKQNLGRL-----NDIKIGAPGELLHTIDIGMLTTPDRDFPAKDKAHR-EYFOTI 318
DB 260 FEDEN--GNLGIIESERIKFSAPSELLIONIDGLMYKPRGNIVIKELERTAVDYFOKV 317
QY 319 PYSRMIVNNAAPHLKVEMLPTGELLTMDPONGGWHSGTHQRIGKELVSHGIDNANYG 378
DB 318 PYSKLIFFSDYTFHFEKTIPLNGSVYVTEKSDIGGWHOGDMREAVGKALVSTGINNANLG 377
QY 379 LNSTAGLGENSEHPYVVAQAAHNSRGNANGITQVHGSGGGGIVTLDSTLGNFESHEVGH 438
DB 378 IVASSGYSQ-QYNRLTNHTAHTNIGYVNGVYVHGSGGGGIVTLENTLNENSHELGH 436
QY 439 NYGLGHVYDGFSGVHRSARENNTGWGDKRFLIPNEYPSOTNEKSL--NNOCQEPF 496
DB 437 NYGLGHVYAG--GTSH---GPDTSWGDYTKREFIANFDMKRSQSNIRPDNQEVKPF 490
QY 497 -DGKKGFGDAMAGSGPF--SAANRFTWYTPNSAIIORFENKAVDFSRSSGFSKWNAD 553
DB 491 MDXYTLMDAMSGGYDHQNGIISRYTLHPHYVARIIQDLKNGAVV--INNDYVWD-E 546
QY 554 TQCEPYEHTIDRAEQITASVNELSESKMAELMAEYAVYVHVMNGNTRNIYIPTASAD 613
DB 547 LKNIYVYKGT-----NEKVPKKGVPVVTILGVYDPDKIN-----PSQLYPPTYS-- 591
QY 614 NRGSL-----TINHEAGYNSY---LFINGDEKVSQ----- 642
DB 592 NYGNIFDLEKPRSESLKGQVYKOVNYLDNRVNTHHMTLVNKRKEKICRFSLYSPKGGK 651
QY 643 ---GYKK-----SFVSDQFVKERDVDTREARKPEQVPTTLVG-----Y 682
DB 652 FEFGLVEDIENKICTGSRSHIYLEDG---KKNP-----ESKYNDYF---LLSIDGGEIS 701
QY 683 YDPEGL-----SSYIYAMYGA-----YGFYVSDSQNLS--- 713
DB 702 YVPDSTIGESKICSLKMSGTVYGAGFIGKNSCRQIDGVFMNGFQWAFITLQSGVNSTYW 761
QY 714 DNDQLOVDYTKGQLR-FRLANHRANTVYKPHINVPESQPTQATLVCKNKLDTKSL 772
DB 762 SNECVLKIKDDNNIESISIPNTRIEKNSKNLHLNISREKPIIDINIVYCGSHELTISIKV 821
QY 773 TPAPEGLTYTVNGOALPAKNEGICVSVNSGKRYCLPVGQSGYSLPDWIVGQEVYVDSG 832
DB 822 SDNPD-----IKLLKGPIL-----VQGEHCYT-----SYEPKLPSSG 852
QY 833 AKAVILSDMDNLG-YNRIGEFVGNVNPADMKVKKAWNGOYL---DPS-KPMSRMV 884
DB 853 -----WFKHYDNFEPKNEINHELKGRVND-----NDEYICRFNFSDSDRMKFV 897

RESULT 4

O9KN18
ID O9KN18 PRELIMINARY; PRT: 1335 AA.
AC O9KN18;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Taga-related protein.
GN VCA0148.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=566;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=El Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae";
RL Nature 406:477-483(2000).
DR EMBL: AE004356; AAF96061.1; -;
DR TIGR: VCA0148; -;
KW Complete proteome.
SQ SEQUENCE 1335 AA; 149521 MW; 0B339F0DBC2DEFD CRC64;
Query Match 14.2%; Score 671; DB 16; Length 1335;
Best Local Similarity 25.3%; Pred. No. 4.6e-35;
Matches 231; Conservative 108; Mismatches 248; Indels 318; Gaps 36;
QY 39 NDLOGSLAAEVKFAQSOILPAHPK--EGDSQPHLTSL--RKSLLVVPYKADKTPVQVE 94
DB 353 NHLOGSLEGLSITQHTSVK-PKGNELTGQCHLDAIMNREALLLFTPOQGEINOVRAE 411
QY 95 ARDDNNKILGTLYPSSLPDTIYHLDGVPEGGIDFTPHNCTKKLIINTVAEYVKLSLAS 154
DB 412 VFLDGLVQVITLML-PPSAL-----AASQPEKGRMKV----- 443
QY 155 GSSTHSLTNALVEIHTANGRWVRDIYLPQGDLEGKMYRFVSAGYSSTVYGDRTV 214
DB 444 ---VFSLA----- 449
QY 215 LSVGNITLLFYVNGWFRSGELENRRITYAQHINSAELPAHWIVPGLNIV-----KQGN 269
DB 450 -----WSLPQMDWMMKPGLSURLTDLNLCREGV 476
QY 270 LSGRLNDIKIGAPGELLHTIDIGMLTTPDRDFPAKDKAHR-EYFOTIYPSRMIVN 328
DB 477 LS--QGEIQFGAPVELVIONIDIGMLMPDRDNTMIONLPTLAADYFQKIPASKLYADY 534
QY 329 APLHKVEMLPTGELLTMDPONGGWHSGTHQRIGKELVSHGIDNANYGLNSTAGLGN 388
DB 535 TPAHFPVYVTPNGVYVYTKSGASTGGWHSQDMREAIKAKVYSTGINNANVGVSSAGYSQ 594
QY 389 -----SHPYVVAQAAHNSRGNAN-----GLOVHGSGGGGIVTLDSTLGNFESHEV 437
DB 595 YNRRFNH-----ITAHNTVGIYTKKDTDLPOVYVHGSGGGGIVTLEATGNWSEHLG 648
QY 438 HNYGLGHVYDGFSGVHRSARENNTGWGDKRFLIPNEF-----YPSQTEKSKLNQ 491
DB 649 HNYGLGHV--PYMASIH---DLESQGWDAFHQRFIGNLHMKGDVYVYTOQGD-----D 696
QY 492 CQEPF-DGHPFGFDMAGGSP--FSAANRFTWYTPNSAIIORFENKAVDFSRSTGFS 548
DB 697 IVPPEKDAFRDLDAQNGEQEYVGTISRFTLEHPAQSRKAQRMMNGNFGNLSHSPGYV 756
QY 549 KWNADTQCEPYEHTIDRAEQI-----TASVNELSES----- 580

```
Db 757 QWQETQRYKAVETDTPKPOGVVPTLLGIYDQNGEQEYVGTISRFTLEHPAQSRKAQRW 816
QY 581 KHAELMAEYAVVYKVMWNGNTRNIYIPTASADNRGSGILTIHAEAGYSY--LFINGDEK 638
Db 817 EOGEOYLE-----GW-----QAAGD-----LT-QAEIOYNOWOTLLIDCQQL 852
QY 639 VV-----SOGYKKSFSVSDQFQKRDVVD-TREARKPEQFGVPTVTLVGYDPEGTLS 690
Db 853 PICRFDYNTNGQSATFY--GSLNAQRNVEGSRDMR-----WYNDYDIDSPVG 899
QY 691 SYIYPAMYAGCYFTYSDDSQ-----NLSND----- 716
Db 900 QYELLSQFGAGNVTYTPNAEIGEYOLCTLNKPHNNGSHDGAGFVRNGRCQEOGVKXNAE 959
QY 717 -----COLQVDTKEGOLREFLANHRANNTYMKKFIHNPTE- 752
Db 960 GRWVYATRNSEVLSRTLASORCLVEYHRNGSTHALDGNRKHSKESKRFVNLMSMEK 1019
QY 753 SQPTQATLVCKNKILDTKSLTPAPEGLTYVNGQALPAKENEGCIVSVNSGKRYCLPVGQ 812
Db 1020 GVPQTQVSLSCSD-LNGTSTLT-----RFTPD-QNPPLDKLKGPII-----IGQ 1060
QY 813 RSGYS 817
Db 1061 EYGYS 1065

RESULT 5
Q8D5P4 ID Q8D5P4 PRELIMINARY: PRT: 875 AA.
AC Q8D5P4;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DE 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Conserved hypothetical protein.
GN W20866..
OS Vibrio vulnificus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=672;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CMCP6;
RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
RA Choy H.E.;
RT "Complete genome sequence of Vibrio vulnificus CMCP6."
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF016811; AAO07787.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 875 AA; 97611 MW; ABE604A7979BC96 CRC64;

Query Match 12.1%; Score 571; DB 16; Length 875;
Best Local Similarity 27.3%; Pred. No. 8.8e-29;
Matches 182; Conservative 99; Mismatches 215; Indels 170; Gaps 27;

QY 259 PGLNLVI-----KGNLSGRNDIKIGAPGELLHTIDIGMLTTPRORFDAKO-KEAHR 312
Db 3 PGLSLRLTDNLGRGVLIS--QGEIYFGAGPELVIONIDMGMLIEPRNTHMIQMLPTLAA 60
QY 313 EYFQIPVSRMIVNNYAPLHKVEMLPTEGELLTDMDPGNGGWHSGTMRQIRIGKELVSHGI 372
Db 61 DYFQKIPASKLVNADYTPAHFPVITPMPGVVYTDKASGTGGWHSGDMREAIKGAWVSTGI 120
QY 373 DNANYCLNSTAGLGEN-----SHPTVYVAQLAANRSGNYAN-----GIQVHGSGGGGI 421
Db 121 NNANVGLTSAGYSQQYNRRFNH-----ITAHTNVGYITTKDTLDLPQIVYHGGSGGGGI 174
QY 422 VTLDSTLGNFEFSHEVGHNYGLGHVYDGFSGVHRSAAENNNSTGWGDKKRFIPNF----- 477
Db 175 VTLEATTGNWSEHGLHNIGLGRW--PYMASIH-----DNESGWGWDAPHQFIGNLHWKG 228
QY 478 --YPSOTNKSCLNNOCQEPF-DGHKFGFDAMAGGSP--FSAANRFTMYTPNSSAIORE 532
Db 13 LSLSSNVQVQATTNEABGCIISRLNGEKYCLKRYGERSYSLPSWIYHPDVPQAPSGVSM 72

Db 229 DVYTOQGD-----DIVPPKDAFRELVDQAQNGEQEYVGTISRFTLEHPAQSRKAQRW 282
QY 533 FENKAVFQSRSTGFSKVNADTQEMEPYHTIDRAEQITASYNEL-----SESKMA 583
Db 283 MNNGFNLDSSSPSGYVOWDQAARQYQTVETDTPKPOOTGVAVMTLLGIYDYPYENPQI- 341
QY 584 ELMAEYAVVYKVMWNGNTRNIYIPTA-----SADNRGSGILTIHAEAGY 627
Db 342 -----YPLVYNGVNFELPGQVOCAGPEQGVADLTFAELSDSQTWOTLRMDGEOQRV 396
QY 628 NSYLF--INGDEKVSQGYKKSFSVSDQFQKRDVVD-TREARKPEQFGV--PVTTLVGY 683
Db 397 CRFTFOAAGDSAVFVGVDQS-----TDKCSSGRDLQWHINSNMTSAQGDY 443
QY 684 DEPGTLLSYVIYPAH-----YCAVGFTYSDSDSONLS----- 713
Db 444 E-----LLSYGRGAVYITTPPEVGEVTLCTLNKSGTDHAGGFVVGNNCEQISGVNKHKG 500
QY 714 -----DNDQCLQVDTKEG--QLRFLANHRANNTYMKKFIHNPTE 752
Db 501 TWRYALRGDEVLRPSYQYQGGCOLDVEFANGCASERVVLNASHRSVND--NKPVNLA 558
QY 753 S-OPTQATLVCKNKILDTKSLTPAPEGLTYVNGQALPAKENEGCIVSVNSGKRYCLPVG 811
Db 559 NGVPTQVRLSCSDRNGETELTRFTPE-----QNPPIADLKGPII-----IG 599
QY 812 ORSGYS 817
Db 600 OEYGIS 605

RESULT 6
Q9KTH2 ID Q9KTH2 PRELIMINARY: PRT: 957 AA.
AC Q9KTH2;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DE 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Hemolysin-related protein.
GN VC0930.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-El Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae."
RL Nature 406:477-483(2000).
DR EMBL: AE004176; AAF4092.1; -.
DR TIGR: VC0930; -.
DR InterPro: IPR002048; EF-hand.
DR PROSITE: PS00018; EF_HAND; 1.
KW Complete proteome.
SQ SEQUENCE 957 AA; 103989 MW; 4C6B40FBC755CDAD CRC64;

Query Match 6.2%; Score 290; DB 16; Length 957;
Best Local Similarity 53.8%; Pred. No. 2.9e-10;
Matches 57; Conservative 14; Mismatches 35; Indels 0; Gaps 0;

QY 779 LHYTVNGQALPAKENEGCIVSVNSGKRYCLPVGQSGYSLPDWIVGQEVYDQSAKAKVL 838
Db 13 LSLSSNVQVQATTNEABGCIISRLNGEKYCLKRYGERSYSLPSWIYHPDVPQAPSGVSM 72
```


[illegible][illegible]

Db	410	-----RKWNWNIQ	-----DAFDKGI	VGFWNDECDEN	YFNGFGNMMNERA	450	
Qy	446	-VDGFKGSVHRS	AEENNNTWG	-----WDGDKR	FIPIFNFP	485	
Db	451	ITDG	-----QRAYK	QRVWLSLRNY	YAGAORYCY	500	
Qy	486	-SCLNNOCQEP	PDGHKFGFD	-----FSAANR	ETMYTPNS	538	
Db	501	LSAVN	-----LGEAK	WMDTGGF	NDGDPPT	550	
Qy	539	-----FDSRS	TGFSKWNAD	QEMPEPYE	HTIDRAEQ	581	
Db	551	RYPWAFG	STAEAAAK	KAKOLRY	TLPIYISY	610	
Qy	582	MAELMAEY	AVVWVHM	WGNTNRIY	IP-----TASAD	630	
Db	611	EAWMFC	DYMLVSP	VVNVQ	GTSKSIY	665	
Qy	631	LFTNG	-----DEK	VVSOGYK	KSFVSD	656	
Db	670	LFIKSG	AIPTQ	QFENVY	GEKKITD	VYVDAF	729
Qy	659	DVVDTE	AREAK	RFQGF	VPTVLGY	YDPSGT	695
Db	730	MTLQ	STDSK	SVQFN	IDKNT--	-GSTPD--	705
Qy	696	AMYG	AYGFTY	SDDSON	LSNDCQ	LQVDTK	755
Db	786	ALKS	ASGEGY	ASGT	-----DTY	GNVYIK	826
Qy	756	TQAT	LVCKNK	-----I	LDPKSL	TAPA	808
Db	827	VTTIA	ANPAG	TYTG	PVTS	LASK	875
Qy	809	PV	810				
Db	876	PI	877				
RESULT 8							
QD	Q45780	PRELIMINARY;	PRT;	1038	AA.		
AC	Q45780;						
DT	01-NOV-1996	(TEMBLrel. 01, Created)					
DT	01-NOV-1996	(TEMBLrel. 01, Last sequence update)					
DT	01-MAR-2003	(TEMBLrel. 23, Last annotation update)					
DE	Outer membrane protein.						
GN	SUSC.						
OS	Bacteroides thetaiotaomicron.						
OC	Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;						
OC	Bacteroidaceae; Bacteroides.						
OX	NCBI_TaxID=818;						
RN	[1]						
RP	SEQUENCE FROM N.A.						
RC	STRAIN-BT5482;						
RX	MEDLINE=96146534; PubMed=8550519;						
RA	Reeves A.R., D'Elia J.N., Frias J., Salyers A.A.;						
RT	"A Bacteroides thetaiotaomicron outer membrane protein that is						
RT	essential for utilization of maltooligosaccharides and starch.;"						
EL	J. Bacteriol. 178:823-830(1996).						
CC	-1- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).						
DR	EMBL; L49338; AAA95938.1; -						
DR	InterPro; IPR001395; Aldo/ket_red.						
DR	InterPro; IPR000531; TonB_box.						
DR	Pfam; PF00593; TonB_dep.Rec.1						
DR	PROSITE; PS00063; ALDOKETO_REDUCTASE_3; 1.						
KW	Membrane; Outer membrane; Receptor; TonB box.						
SQ	SEQUENCE 1038 AA; 115886 MW; DD46CE38A2DC5E1 CRC64;						

Query Match 3.4%; Score 161; DB 2; Length 1038;
Best Local Similarity 18.1%; Pred. No. 0.098;
Matches 180; Conservative 120; Mismatches 328; Indels 366; Gaps 46;

QY 39 NDLOGSLAEVFAQSQILPAHPKEDGDSOPHLTSLRKSILLVR-----PVKADDTKP---V 91
 Db 131 NDMTGSVMA-----IKPDELSG-----ITTAQDMLSKGIAGSVSVIPGCGGA 177
 QY 92 QVEARDNNKILGTLTLYPPSSLPOTIYHLCQVPEGGIDFTPHNGTKKIINTVAAEVNK-- 149
 Db 178 QIRIRGSSU-----NASNOLIVIDGL---AID---NEGIKGMANGLSMVNPD 221
 QY 150 -----LSDAGSSSHSHLTNNALVEIHTANGRWTRDYLPQGPDEKGRVFFVSSAGYS 203
 Db 222 IETLTVLKASATAIYGSASNGVITITTKKG-----NGQAPSVTYNGSVSFS 270
 QY 204 STVFYGDRTVLSVGNLTFLKVVNGQWPSGELNNRITYAOHIWSAELPA-----HW 256
 Db 271 KT-----QKRYDVLSDGEYA-----YANQMGDKLPADLGTANTDW 307
 QY 257 -----IVPGLNLVIRG--NLGRANDIKIGAPGELLHTIDIGMLTTPRDRFPAK 306
 Db 308 QDQIFRTAVSTOHVHSINGFKNLPRVRS-----LGYTDONGIVKTSNER-RETA 356
 QY 307 DREAHEXYEQ---TIPVSRMIV---NNYAPLHLKE---VMLPTGELLTMDP-----GNGG 353
 Db 357 SVNLAFFEDHLKFNINAKFMKNGKRYADSCRIRALAIIDTPRVSNEPDYFTGTY 416
 QY 354 WHSGTMRQIRIGELVSHGIDNANYGLNSTAGLGE-----NSHPYVVAOLAAHNSRG 404
 Db 417 WQN-----INSTGFSNPDWKYTSNPNSPONPLAALCLKNDKG 454
 QY 405 NYANGIQVHGGSGGGIVTLTDLSTGLNEFSEHVECHNYGL-----CHYVDGPKG 451
 Db 455 N-----SNDFCYGVNDVYKPHFLPDLRLHASIGEYAEQT 491
 QY 452 SVHRSANNSTWGDG---KKREIPNFPYSQTNKESCLNQCQEPDGHKGFEDAMA 507
 Db 492 TVVSPYSGNNGYWGNDVTOYKYNLSYNYVQYI--KSLGAND-----FDIMV 538
 QY 508 GGSPPSAANRFTMYTPNSSAIIQRFENKAVPDSRSSGTFSK---WNADQOMEPYEHTI 564
 Db 539 GSEE-----QHFRN-----CFEEOGQWDSYTOE---PHDA 566
 QY 565 DRAEQITASVNELSESKMA---ELMAEYAVVVMHMGNCWTRNIYIPTASADNRGSLT 620
 Db 567 KLRQATAYATRNLTLSYFGRLYSLNRLYLTFTWRWOGS-----SRFSKDNRM--- 615
 QY 621 INHEAGYNSYLEFNGDEKVVSGYKSFYSQGFWKERDVVDVTREARKPEQFGVPVTTLV 680
 Db 616 -----GTFPSLALGWK---IKEENFLKDVNVLSDLKL---LWGIGITGOQ 654
 QY 681 GYDPEGTLSYIYPAMYGAYGF--TYSDDS--QNLSDNDQQLQVDYTKEGQLRFLANHR 736
 Db 655 NIGDDFAYLPLXYVNNYEAQYFPGDTYTSRPAFNEKLEKERTTTWAGLDGFLNGR 714
 QY 737 ANNTVNKFERINVPTEOQTATLVNCKNKLDTKSLTPAPEGLTY-----TVNGQ 786
 Db 715 ITGGIDGYF-----RKTMTCVTALRSPMNLQCPDDEYRFTYCKLRYGFSINAK 763
 QY 787 ALPAKE-----NFGCIVSVNSG--KRYCLPVG----- 811
 Db 764 PIVTKDFTWDSYNITWNHNEITKLGTGGDSYIYVEAGDKISKGNNTKVOAHKVYAANS 823
 QY 812 -----QRSYSLPWHVCOEYVYDSGAKAKYLLSDWNLSVNRIGETFGVGNV 858
 Db 824 FYVSRGNNTKVOAHKVYAANSFYVYQYVDENGKPI-----ENMEFYDRNGN--GTID 874
 QY 859 PADMKVKAWSG-----CYLDFSKPSRMR 882
 Db 875 SGDKYIYKPKAGDVLMLGTLKSKQYKNKFDPSFSLR 908

RESULT 9
 Q46034
 ID Q46034

PRELIMINARY; PRT; 2367 AA.

AC Q46034;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Toxin B.
 OS Clostridium difficile.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium
 OC Clostridium
 OX NCBI_Taxid-1496;
 RN [1]
 SEQUENCE FROM N.A.
 RC STRAIN-isolate 1470;
 RX MEDLINE=96079281; PubMed=7494480;
 RA von Eichel-Streiber C., Meyer zu Heringdorf D., Habermann E.,
 RA Sattigen S.;
 RT *Closing in on the toxic domain through analysis of a variant
 RT Clostridium difficile cytotoxin B*;
 RL Mol. Microbiol. 17:313-321(1995).
 DR EMBL; 223277; CAAB0815.1;
 DR InterPro; IPR002479; CW_binding.
 DR InterPro; IPR001950; TIF_SUII.
 DR Pfam; PF01473; CW_binding_1; 18.
 DR PROSITE; PS01118; SUII_1; 1.
 SQ SEQUENCE 2367 AA; 269186 MW; EF9823DAE70427F3 CRC64;
 Query Match 3.4%; Score 159.5; DB 2; Length 2367;
 Best Local Similarity 18.4%; Pred. No. 0.45;
 Matches 185; Conservative 149; Mismatches 321; Indels 353; Gaps 49;
 QY 27 NSATYFNTSOPINDQGLAAEVFAQSQILPAHPKEDGDSOPHLTSLRKSILLVRPVKAD 86
 Db 1454 NSELQKNIPYSFVDSGK-----ENGFIGSTKEG-----LFVSELPDVLISKVYMD 1501
 QY 87 DKTP-----VOYEARDNNKILG-----TLTLYPPSSLPDIYHLDGV 124
 Db 1502 DSKPSFGYGNLKDVAIVITKDNVILTYLKKDDIKISLSLTQDEKTIKLNSVHLD-- 1559
 QY 125 PEGGIDFTPHNGTKKIINTVAAEVNKLSDSAGSSSHSHLTNNALVEIHTANGRWTRDIYLP 184
 Db 1560 -ESG-----VAETLKFNKRGS---TNTSDLSMFLSMN---IKSIFV- 1596
 QY 185 QGPDEKGRVFFVSSAGY--SSIVFYG-----DRKVTLSVNGT 220
 Db 1597 ---NFLQSNKIFLDANFIISGTSIGQFEFCDENNNIQPIFKENTLTETNLTLYVGNR 1653
 QY 221 L-LFKYVNGQWFRSGELENNRITYAQ-HWSAELPAHWIYPGLNLVILKGNLSG----- 272
 Db 1654 QNMIVEPNYDLDDSGDISSTVINFSQKLYGIDSCVKNKVISPNYITDEINITYETNN 1713
 QY 273 -----RLNDIKI-----GAPGELLHTIDIGMLTTPRDRF-DFAK 306
 Db 1714 TYPEVIVLDANYINEKINYNINDLSIRYVNSDNGDNDFILMSTSEENKVSQVIRFVNVFK 1773
 QY 307 DKE-AHREYF---QIPVSRMIVNNYAPLHLKEVNL----- 338
 Db 1774 DKTLANKLSFNFSQDQVPVSEIIL-SFTPSYEDGLIGYDGLGLVSLYNEKFIYNNFGMM 1832
 QY 339 -----PTGELLTD-----MDPFGNGHSGTMRQIRIGELVSHGI 372
 Db 1833 VSGLIYINDSLIYFKPPVNNLITGFTVVGDDKYFNPINGGAAS-----IGETI---I 1882
 QY 373 DNANYGLNSTAGLNSHPYVVAOLAAHNSRGNY-----ANGIQVHGGSGGGIVTLDS 427
 Db 1883 DDKNYTFNQS-----VLQTVGFSTEDGFKYFAPAN-----TLDEN 1918
 QY 428 LGNE---FSHEVGHNYGLGHYVDGFKGSHVRSANNSTWGW---DGDKKRFPNFPSPQ 481
 Db 1919 LEGAIDFTGKLIIDENIYYFEDNYRGAVE-----WKELGEMHYFSP----- 1962
 QY 482 TNEKSCNLNQCQEPDGHKGFEDAMAGGSPFSAANRF--TMYTPNSSAIIQOREF---EN 535
 Db 1963 -----TGKAFKGLNLQIGDDKYFNSDGVNKGQFVSINDN 1996

[illegible]

RC STRAIN-ATCC 824 / DSM 792 / VEM B-1787;
RX MEDLINE-21359325; PubMed-11466286;
RA Noelling J., Breton G., Omeichenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum";
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL: AEC007786; AAK80833.1;
DR InterPro: IPR005085; CBM_25.
DR InterPro: IPR000322; Glyco_hydro_31.
DR InterPro: IPR000531; TonB_boxC.
DR Pfam: PF03423; CBM_25; 1.
DR Pfam: PF01055; Glyco_hydro_31; 1.
DR PROSITE: PS00430; TONB_DEPENDENT_REC_1; 1.
DR Hydrolase; Complete proteome.
SQ SEQUENCE 1157 AA; 129411 MW; 1DB1728D833F7578 CRC64;

Query Match 3.3%; Score 156.5; DB 16; Length 1157;
Best Local Similarity 19.3%; Pred. No. 0.23;
Matches 198; Conservative 120; Mismatches 335; Indels 371; Gaps 56;

QY 83 VKADKTPVOVEARDNNK-----ILGTLTLPSSLPDIYHLCQVPEGIDFTPHNCTK 138
DB 35 VRADTKVNVKNKSSQSKFKAKLNGTLKIKGKDETIIRICEPQVFKVDYKP-NGKS 93
QY 139 KIINTVAEYNKLSDSGSIHSLTNALVEIHTANGWRDI-----YLPQGPDL 189
DB 94 SKDTLVVDPNK-KWSTGNIVSDIISDPHV-----ITTKMKVLKINWEDLSILVY-----DL 144
QY 190 EGMWRFVSAGYSSTVF---YCDKRVTLVSGNTLLFKYVNGQWFR-----SGELEN-NRI 241
DB 145 OGRLLKQDSTASKTASFTHNSGDR-----FYGINGYFQEDSSKGLMRNGTES 193
QY 242 TYAQHTWSAELPAHVIVPGLNVIKGNLSGRND-----IKIGAPGEL 285
DB 194 VTAGYGHGCGSPFVWNSDGLLVSDSGSFTIGTSLKYDGIKTDYDYYVMGNPKREI 253
QY 286 LLHTIDIG-----MLTTPDRF-----DFADKEAHREY-FQTIPVSRMI 324
DB 254 LSESDVSGKAPFPKAWNGFTNTQWGNLSLGGTNDKALKLSVINTYRSKOLPIDNFC 313
QY 325 V-----NNYAPLHLKEYMLP---TCELLTMDPGNGHSGTMRQRI-----GKE 366
DB 314 LDFDKWKWGDNYGEFKWNTNFPDSQGLKAYMD-SKGLKMTGIMKPRILADSQGRY 372
QY 367 LVSHG-----IDNANYGLNSTAGLGENSEHPYVVAQLAAHNSRGNANG 409
DB 373 VTSKGWMLPCDSASDYCSCKMKNVNFAL-----PQV-----RWKWN 412
QY 410 IQVHGGSGGGIVTLDSTLGNESHEVGHNYGLGHY-----VDGFKGSHRSANN 460
DB 413 IQ---GAFDKGIVGF-----WNDCDENVNFNGNMMERAIYDG-----ORRHKN 456
QY 461 NSTWG-----WDGDKRFTIPNFPSTQNEK-----SCLNQCQEPFDGKH 500
DB 457 QRWVSLNRNYAGAQRYSGYGMNSGD-----ISTGDSMAQRERMLSAVN-----LGEAK 506
QY 501 FGFDD---AMAGGSP-----FSAANFTWYTPNSSAIIQRF--ENKA-----VFDRSSTGFS 548
DB 507 WGMTDGGFGNDPTPENYARWMEFSAFTP-----IPRVHGQDNKYRYPWAFGSTAEATAK 561
QY 549 KWNADTOEHPYEHITDRAQITA-----SVNELSEKMAELMAYAVYKVTM 596
DB 562 KAMQLRTPITYISYDRSASQSLGLVRSLSMEXYPNDSNAANDKAWMFGDYMLVSPV 621
QY 597 WNGWTRNIVIP-----TASADNRGSIILTNHEAGYNSY---LFING----- 635
DB 622 QEGQTSKSIITLPGNWDYTTGREYTG--QTINAVDSTNWDIPUFIKSGAIITQDFE 680
QY 636 ---DEKVVSGYKKSF-----VSDGQFWKERDVIDTREARKPEQFG 673
DB 636 ---DEKVVSGYKKSF-----VSDGQFWKERDVIDTREARKPEQFG 673

DB 681 NYVGEKKITDYVDYDAPPNGNEASSFDYDDGTSYNYENGSYFDQKMTLERAKDLKSVQFN 740
QY 674 VPVTTLVGYDPEGLTSSYIP-----AMYGAYGFTYSDDSQNLSDNDC 717
DB 741 ISPTK--GY--KSDLKNYIVKMHVKSQGDVTVGGRITRYASY-----DELKNAQGEY 791
QY 718 OLOVDTKEGQLAPRL-ANHRANNTVYKHFHINVPESQPTQATLVCKNKIL-----DTKSL 772
DB 792 VGTDTYGSVYIKYSAGHDKN-----INP-----CNQVLTAYADVK-- 830
QY 773 TPAPEGLTYVANGQALPAKEN-----EGCIVSVNSGKRYCLPVQSRSGYSLPDI 822
DB 831 -----GGTYT-SPQKVSLSKASDPNAAIYVTLDTAPTAVNSTK-----YTGP--- 870
QY 823 VGOEYTVDSGAKAKVLLSDWNLNSYRIGEFVGNVNPADKKVKAANGQYLDFSKPRSMR 882
DB 871 -----ITIDSKTLKFIVRD-----ANGNESDV-----FTEQYTTY-----IK 903
QY 883 VYK 886
DB 904 VHK 907

RESULT 12
QYF931 PRELIMINARY; PRT; 2367 AA.
ID QYF931 AC QYF931;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DB Cytotoxin B.
CN TCDB.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID:1496;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-5340;
RX MEDLINE-20448897; PubMed-10992443;
RA Sambol S.P., Merrigan M.M., Lierly D., Gerding D.N., Johnson S.;
RT "Toxin gene analysis of a variant strain of clostridium difficile that
RT causes human clinical disease";
RL Infect. Immun. 68:5480-5487(2000).
DR EMBL: AF217292; AAG18011.1;
DR InterPro: IPR002479; CW_binding.
DR InterPro: IPR001950; TIF_SUII.
DR Pfam: PF01473; CW_binding.1; 18.
DR PROSITE: PS01118; SUII.1; 1.
SQ SEQUENCE 2367 AA; 269337 MW; D5EE715E5BD41E2F CRC64;

Query Match 3.3%; Score 154.5; DB 2; Length 2367;
Best Local Similarity 18.4%; Pred. No. 0.95;
Matches 185; Conservative 148; Mismatches 322; Indels 353; Gaps 49;

QY 27 NSAIYFNTSQPINDLQSLAAEVKFAQSQILPAHPKQSGSQPHLSLRKSLSLVPRVKAD 86
DB 1454 NSELQRNIPSYFVDSGK-----ENGFGSTKEG-----LFSVSELPDVVLSKVYND 1501
QY 87 DKTP-----VQVEARDNNKILG-----TLTLPSSLPDIYHLDGV 124
DB 1502 DSKPSFCYYSNNLKDVKITKQNVNLTGYYLKDDIKISLTLQDEKTIKLSVHLD-- 1559
QY 125 PEGGIDTPPINGTKKINTVAEYVKNKLSADSGSIHSLTNALVEIHTANGWRDIYLP 184
DB 1560 -ESC-----VAETLKFPNRRKS---TNTSDLSMFSLESNM---IKSIFV- 1596
QY 185 QGPDLECKMVRFYSSAGY---SSTVFYG-----DRKVTLSVGWT 220
DB 1597 ---NFLQSNKIFLDANFIISGTSIGQFEFCDENNNOPIFKNTLTNTLYVGNR 1653
QY 221 L-LFKYVNGOWFRSGELENRIYTAQ--HWSAELPAHVIVPGLNVLKQNLGSG----- 272

1654	QNMVPEPNDLDSGDISTYVNF	SKYLYGDS	CVNKVVIS	PNIS	NYTDD	INTPTV	ETNN	1713																																														
273	-----RLNDIKI-----	GAPCELL	LHTID	GLMTTP	DRDF	-DFAK	306																																															
1714	TYPEVIL	DANYINEK	INVIN	DLSIRY	WSDG	NDFILMST	SEENKV	SOVAKIREVNVFK	1773																																													
307	DRE-AHREYF-----	QTI	PVSR	MI	VNYAPL	HLKEVRL	-----	338																																														
1774	DKTLANKL	SNFSDKQD	VPVSEIIL-SFT	PSYED	GLGYD	GLVSLY	NEKYVIN	NFGMM	1832																																													
339	-----PTCELLTD-----	MDP	GMG	WHS	GTM	RORIKEL	VSHOI	372																																														
1833	VSGLI	YINDSLY	FKP	PVNN	LITG	FTVGD	KYFNP	INGAAS	-----	ICETI	1882																																											
373	DNANY	GLNSTAG	GENSH	PPV	VAQ	LA	HN	SRGNY	-----	ANG	IQVH	GSGGO	IVTLD	ST	427																																							
1883	DDKNY	FNOSG	-----	VLQ	TGF	STED	GFKY	FAPAN	-----	TLDEN	1918																																											
428	LGNE	-----	FSHEV	GHNT	GLHY	DG	FGK	SVH	RS	AE	NN	STWGV	-----	DDK	KRF	IPNE	PSQ	481																																				
1919	LEGE	ADFT	CKL	LIDEN	YIF	DNY	GAVE	-----	WKE	LO	GE	HHY	FPE	-----	1962																																							
482	TNBK	SLC	LN	OC	PP	DG	HK	FG	DAM	AG	GS	FF	SA	AN	RF	--	TM	YTP	NS	AI	IQ	RF	--	EN	535																													
1963	-----TGKAF	GL	NO	IG	DD	KY	FN	SD	GV	M	K	FV	S	IND	1996																																							
536	KAF	DSRS	-----	NA	TO	E	ME	P	Y	E	H	T	I	D	R	A	E	O	I	T	A	S	V	N	E	L	S	R	K	M	A	E	L	M	587																			
1997	KH	F	D	S	G	V	M	K	V	G	T	E	D	G	H	F	T	A	F	A	N	G	E	M	O	I	G	V	F	W	T	E	D	G	F	K	F	A	H	N	E	D	L	G	E	E	E	I	S	2056				
588	EY	A	V	K	V	H	M	N	G	W	N	T	R	I	Y	I	P	T	A	S	A	N	R	G	S	I	L	T	I	N	H	E	A	G	Y	N	S	Y	L	F	I	N	G	D	E	K	V	S	Q	Y	K	S	647	
2057	YSGIL	-----	NFN	K	I	Y	-----	F	D	S	T	A	V	M	K	L	E	D	G	S	K	I	Y	-----	D	B	T	A	E	A	V	I	G	L	S	2103																		
648	FVSD	G	Q	F	M	K	E	R	D	V	D	T	E	A	R	K	P	O	F	G	-	V	P	T	L	V	G	Y	D	P	S	E	T	L	S	S	I	Y	P	A	M	Y	C	A	G	E	T	Y	S	706				
2104	LIND	Q	T	F	N	D	O	G	I	-----	M	O	V	G	E	T	I	N	D	K	V	I	F	S	D	S	G	I	E	S	-----	G	V	O	N	I	2145																	
707	DD	S	O	N	L	S	D	N	C	O	L	V	-----	D	T	K	E	G	O	L	R	F	L	A	N	H	R	A	N	T	I	N	V	M	N	K	F	I	N	V	P	T	E	S	O	P	T	A	T	L	V	C	N	763
2146	DD	N	Y	F	I	D	D	N	G	I	V	O	I	G	V	F	D	T	S	D	G	Y	K	F	A	P	A	-----	N	T	V	N	D	N	I	Y	-----	2183																
764	NK	L	D	T	K	S	L	T	P	A	P	E	-----	G	L	T	V	N	G	O	A	L	P	A	K	E	N	E	C	I	S	V	N	S	G	K	R	Y	C	L	P	V	G	O	R	S	-	G	Y	816				
2184	GQ	A	V	E	S	G	L	V	R	V	E	D	V	Y	F	G	E	T	I	E	T	F	T	W	I	D	M	N	E	-----	S	D	K	I	F	O	P	E	T	K	A	C	K	G	I	2236								
817	SL	P	O	W	I	V	G	O	E	V	Y	V	D	S	G	A	K	A	V	L	L	S	D	W	N	L	S	Y	--	N	R	I	G	E	F	-	V	G	N	V	N	P	A	D	861									
2237	N	L	I	D	D	I	---	K	Y	F	E	K	G	I	M	T	G	L	I	S	F	E	N	N	Y	F	E	N	G	E	M	O	F	G	I	N	E	D	2281															

REF: 13

```

RES001 13
Q8KRR3      PRELIMINARY;      PRT: 3692 AA.
ID  Q8KRR3      AC
AC  Q8KRR3;
DT  01-OCT-2002 (TrEMBLrel. 22, Created)
DT  01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT  01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE  Outer membrane protein.
DE  FusoBacterium nucleatum.
OS  FusoBacterium nucleatum.
OC  Bacteria; FusoBacteria; Fusobacteriales; Fusobacteriaceae;
OC  Fusobacterium.
OX  NCBI_TaxID=851;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=PK1594;
RA  Ma X., Kempf M.J., Jewett A., Park H.-H., Shi W.;
RT  "Cloning and analysis of FusoBacterium nucleatum apoptosis-inducing
RL  genes."
RL  Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AF525505; AAM90995.1; -
DR  InterPro; IPR005546; Autotransporter.
DR  Pfam; PF03797; Autotransporter; 1.

```

Db 532 MDTGGFNQDGTNNPNDLYARMQFSALPVPYRVGNHNNHQRPWFSGTAEASKEA 591

QY 568 EQI-----TASYNELSEKMAELMAEY---AVVK--VHMWN--GNW----- 601

Db 592 IQLRSLIPYMYAYERSAYENGGLVRLQAOYPTDAVKNYTDAMWFGDWLLAAPVVK 651

QY 602 ---TRNIYIPTASADN--RGSILT-----INHEAGYNSYLFINGDERKVSQYKKS 648

Db 652 QOTSNDIYLPSSGWDIYARGNAITGGQTIYRSVNPDTLTDMPLEI-----KKG 700

QY 649 VSDGQFWKERDVIDTREARKPEQFGVPTILVG--YYDPEGLSSYIIPAMYGAIGTYS 706

Db 701 IIPQO--KVQDYVGOASVTSVDVDVFPDTSSTSYDDG-----ASYNVESCTYF 750

QY 707 DDSQNLSDNDQLOVDTRREGQLRFLANHRANNTVMKFNHNVTSQPTQATLVCNKI 766

Db 751 KQMTAQDNG-----SGLSFTLGAKSGSYPALQSYI---VKLHSGAGTSVTNNS- 798

QY 767 LDTKSLTIPAPEGLTYVNGQALPAKNEGCIIVSVNSGKRYCLPVGQRSGYSLPDWIVQGE 826

Db 799 -----AANTSASLEALKKAAGES-----WATGKD 823

QY 827 VYVD 830

Db 824 IYGD 827

RESULT 15

Q912M3 PRELIMINARY; PRT; 2468 AA.

AC Q912M3; DT 01-MAR-2001 (TREMBLrel. 16, Created)

DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)

DE Hypothetical protein PA1874.

GN PA1874.

OS Pseudomonas aeruginosa.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

OC Pseudomonadaceae; Pseudomonas.

OX NCBI_Taxid:287;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-ATCC 15692 / PA01;

EX MEDLINE-20437337; PubMed-10984043;

RA Stover C.K., Pham X.-O.I., Erwin A.L., Mizioch S.D., Warren P.,

RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,

RA Garber R.L., Tolentino E., Westbrock-Wadman S., Yuan Y.,

RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,

RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,

RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;

RT Complete genome sequence of Pseudomonas aeruginosa PA01, an

RT opportunist pathogen.";

RL Nature 406:959-964 (2000).

DR EMBL; AE004613; AAG05263.1; -

DR InterPro; IPR001343; Hemlysn_Ca_bind.

DR InterPro; IPR006162; Ppantyn_attach.

DR PRINTS; PR003113; CABINDNGRPT.

DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 2468 AA; 238414 MW; 13596AFAB2C4B899 CRC64;

Query Match 3.2%; Score 148.5; DB 16; Length 2468;

Best Local Similarity 19.5%; Pred. No. 2.5;

Matches 182; Conservative 90; Mismatches 301; Indels 361; Gaps 45;

QY 33 NTSQPLNDILOGSLAAEVKFAQSQILP-----AHP-----KEGDSQPH----- 69

Db 324 NNETPTVTEAPOTTAPAPATDVQVAPDGSSVTGNAEPATGVDTDGDGQPDPTVVVGP 383

QY 70 -----LTSLRSLILVPRYADTKPTQVYARD-----DNNKIL----- 103

Db 384 GSPEVPLNPLNGEVTIVITDTPAGNCSSTPTVYASAPDFPDAPQVNASNGSLGTAEG 443

RESULT 14

Q8RQ09 PRELIMINARY; PRT; 1284 AA.

AC Q8RQ09; DT 01-JUN-2002 (TREMBLrel. 21, Created)

DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)

DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)

DE 6-glucosyltransferase.

CS 2.

GN Maruta K.;

OS Bacillus globisporus.

OC Bacteria; Firmicutes; Bacillales; Planococcaceae; Sporosarcina.

OX NCBI_Taxid=1459;

SEQUENCE FROM N.A.

RN RC STRAIN-C11;

RA "Cloning and sequencing of the genes encoding cyclic tetrasaccharide-

RT synthesizing enzymes from Bacillus globisporus C11.";

RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB071929; BAB88404.1; -

DR InterPro; IPR005084; CBW_6.

DR InterPro; IPR000322; Glyco_hydro_31.

DR InterPro; IPR001254; Ser_protease_Try.

DR Pfam; PF03422; CBW_6; 2.

DR Pfam; PF01055; Glyco_hydro_31; 1.

DR PROSITE; PS00135; TRYPSIN_SER; 1.

KW Transferase.

SQ SEQUENCE 1284 AA; 139128 MW; 2167B0DE84F42E9C CRC64;

Query Match 3.2%; Score 149; DB 2; Length 1284;

Best Local Similarity 18.8%; Pred. No. 0.84;

Matches 170; Conservative 105; Mismatches 321; Indels 308; Gaps 42;

QY 105 TLTLYPPSSLPDTIYHLQVPEG--GIDFTPHNGTKKIINTVAEVNKLSDASGSSIHSH- 161

Db 54 TLTVDNGAEPDSDLLIVQVNGILKVDYRNSITPSAKTMDPNTKTSVAVGATINTTA 113

QY 162 -----LTNNALVEI-----HTANGWRVDIYLPQGPDEKWKVRFVSSAGYSTVF 207

Db 114 NPMTITTSNMKIEITKNPVRMTVKADGTTL--FWEPSSGGVFGSDGVRLHATGDN--M 168

QY 208 YGDRKVTLSVGNLTLLFKYVNGQWFRSGLENNRITYAQHISAEPLAHMIVPGLNLVIKQ 267

Db 169 YGIRSF-----NAFTSGDGLLRNSSNHAH----- 193

QY 268 GNLSRLNDIKICAPGELLHTIDIGMLTTPDRDFDFAKDEAHREYFQ--TIPVSRMIV 325

Db 194 ---AGEQGD---SGPLIWSAGYGLLVDSGGVPPYDSTTGQMEFYIGCTPPEGRRYA 246

QY 326 NNYAPLHL-----KEVNLPTGELLTMDPGNGWHSGTMR-----Q 361

Db 247 KONVEYIMLGTPKKEIMTDVGE--ITGKPPMLPKWLSGFNFEMTDNQTFFNNVDYTRAK 305

QY 362 RIGKELVSHGIDNANYGLNSTAGLGENSHYVVAQLAAHNSRNGNYANGIOVHG----- 414

Db 306 NIPIDAYAFDYDKKYGENTYGEFAWNTNFPASSTSLKSTMD-AKGIKIGITKPRIV 364

QY 415 -SGGGGGIYV--LDSTLGNFESHEVGHNYGLGHYVDGPKGSHRSAAE---NNNSTWGWG 468

Db 365 TKDASANVTGCTDATTNGYF--YPCHN---EYQDYFIPVTVRSIDPYNANERAFWNH 418

QY 469 DK---KRFIPNEYPSOTNEKS-----CLNNQCOEPPDGHKFGFDAMAGSP--- 511

Db 419 STDALNKGIWGMNDETDKVSSGGALYWFNFTTGHMSOTMYEGGR-----AYTGAQRW 474

QY 512 -----FSAANRF--TMYTP-----NSSAIIQRFPENKAVFDRSRSTGFSKWN 551

Db 475 QTARTYPCQQRVATTLWSGDIGIYQNKGERINWAGMQ---EQRAVMLSSVNNQVKNWG 531

QY 552 ADT-----QEMEPY-----EHTIDRA 567

QY 104 -----CTITLYPPSSLPD-----TIYHLDGV 124
Db : : : : :
444 VTIVITDGNNGPIGOTSADANGNSFTGSQLPDGTVVNVYVARDAAGNSSPATSTIVDGV 503
QY 125 PEGGIDFTPHNGTKKIINTVAEVNK---LSDASGSIHSLTNNALVEIHTFANGRWVRDI 181
Db : : : : :
504 APNAPVVEPSNGSE--LSGTAEPGSSVTLTDGNGNPIGOTTAD-----ANGNW---S 550
QY 182 YLPQGPDLGSKMVRFS--SACYSS-----TVFYGDRKVTLSVGNLTLLFKYVNGOMF 231
Db : : : : :
551 FTPTPLPDGTVVNVYVARDAAGNSPPASVTVDAVAPATPTVDPNSGTTL----- 600
QY 232 RSGELENNRITYAQHIWSAELPAHMIIVPGLNLVIRKQNI.SGRINDIKIKAPCELLLHTID 291
Db : : : : :
601 -SGTAE-----PGSSVTLTDGN-----GNP----- 619
QY 292 IGHMTTPRDFOFKDEKHAHREYFQTIPIVSRMIVNVYAPLHLKEVMLPTGELLTD----- 346
Db : : : : :
620 IGQVTA-----DGSNGMTFTPTPLPNGTVVYNATAD-----PSGNASSPASVTV 664
QY 347 -----MDPGNGWHSGTMRQIRIGKELVSHGIDNANYGLNSTAGLGENSHPHYVVAOL 397
Db : : : : :
665 DAVAPATPVVNSNGTTLSGT-----AEPGATVTLTDG-NGNP--ICQV 705
QY 398 AAHNSRGNYA-----NGIQVH-----GGSGGGGIVTLDS-----TL 428
Db : : : : :
706 TADGS-GNMSFTPTPLPNGTVVYNATADTASNTSAGSSVTVDSVAPATPVINPSNGTTL 764
QY 429 GNEFSHEVGHNTGLHYVDGFKGSHRSAAENNSTMGWDGKKRFTPNFYPS-OTNEKSC 487
Db : : : : :
765 SG--TAEPGSSVTL---TDGNGNPIGQVNTADGSGNW-----SFTPTPLADGTV 808
QY 488 LNNQCOEPFDGHKFGFDAMAGGSPFSAANRFTWYTPNSSAIIORFFENKAVFDSRSSTGF 547
Db : : : : :
809 VNATATDP-----AGNTSGGGSTVDGVAPTITV-----NLSNGSSLGTA- 850
QY 548 SKWNADTOEMEPYEHTIDRAEQITASVNELSESKMAELMAEYAVVYVHMNONGNWTNRNIYI 607
Db : : : : :
851 -----EPGS-----TVILTDGNGNPIAEVTAD-----GSGNWT---YT 880
QY 608 PTASADNRGSILTIHNEAGVNSYLFINGDEKVVQGYKKSFVSDGQFWKERDWDVTREAR 667
Db : : : : :
881 PSTPIANGIVNVYVVOADAAGNS-----SPGASVT-----VDSQAPA 916
QY 668 KPEQFGVPVTVTLVGYDPEGTLLSSYIYPAMYGAGFTYSDDSQNLSDNDCLQVDTKEGQ 727
Db : : : : :
917 APVVPNSNGTTLSGTAEPGATV-----TLTDGNGNPIG---QVTAD-GSGN 958
QY 728 LRFLANHRANNTVMKPHINVPTESQTOATLVCNNKILDTKSLTPAPEGLTYTVNGOA 787
Db : : : : :
959 WSFTPGTPLANGTVWN-----ATASDPTGNTSAPASTTVD--SVAPAAP-VVNPNSNGAE 1009
QY 788 LPAKENEGCIVSVNSGKRYCLPVGQSGYSLPDW 821
Db : : : : :
1010 ISGTAEPGATVTLTDGSGN--PIGQVNTADGSGNW 1041

Search completed: October 8, 2003, 16:27:19
Job time : 54.5173 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 8, 2003, 16:20:57 ; Search time 21.2134 Seconds
(without alignments)
3912.317 Million cell updates/sec

Title: US-10-002-309b-2_COPY_24_886

Perfect score: 4597
Sequence: 1 ADNNSAIFYNTSQPINDLQG.....AWNGYLDKSPRSMRVYK 863

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4597	100.0	898	2 T42131	probable toxR-regu
2	2604.5	56.7	587	2 T00316	toxR-regulated lip
3	689.5	15.0	1002	2 T09438	toxR-activated lip
4	689.5	15.0	1013	2 B82276	ToxR-activated gen
5	671	14.6	1335	2 A82494	TagA-related prote
6	290	6.3	937	2 H82261	hemolysin-related
7	166.5	3.6	1217	2 F97177	alpha-glucosidase
8	161	3.5	1038	2 JC6027	115K outer membran
9	159.5	3.5	2367	2 S70172	toxin B - Clostrid
10	156.5	3.4	1157	2 F97255	fusion of alpha-gl
11	154	3.4	1211	2 S54500	alpha, alpha-trehal
12	153.5	3.3	2366	2 S10317	toxin B - Clostrid
13	148.5	3.2	2488	2 A83412	hypothetical prote
14	148	3.2	1021	2 I40805	collagenase - Clo
15	147	3.2	1386	2 AC1533	surface protein (L
16	146.5	3.2	1377	2 D90538	hypothetical prote
17	139	3.0	1530	2 AH1396	peptidoglycan anch
18	138.5	3.0	1268	2 B99789	hemagglutinin/hemo
19	138.5	3.0	1270	2 E85649	hypothetical prote
20	138	3.0	2894	2 C64474	hypothetical prote
21	136.5	3.0	973	2 B86547	polymorphic outer
22	136.5	3.0	973	2 F72076	polymorphic outer
23	136.5	3.0	995	2 C81593	polymorphic membra
24	136.5	3.0	3705	2 AD0123	probable autotrans
25	136	3.0	2167	2 AF1489	cell wall-associat
26	135.5	2.9	2529	2 B64635	toxin-like outer m
27	135	2.9	1579	2 B91290	probable invas
28	135	2.9	1700	2 G86131	probable invas
29	134.5	2.9	1026	1 TLB774	tail fiber protein

30	133.5	2.9	2660	2 E85822	probable invas
31	133	2.9	4199	2 S76412	hypothetical prote
32	132	2.9	815	2 H83345	ferripyoverdine re
33	132	2.9	2273	2 T09083	hemagglutinin/hemo
34	130.5	2.8	773	2 JE0387	exo-alpha-sialidas
35	130	2.8	813	2 A40601	ferripyoverdine re
36	129.5	2.8	1886	2 T13169	tiggrin - fruit fl
37	129	2.8	1250	2 D91018	hypothetical prote
38	128.5	2.8	1004	2 B25039	outer cell wall pr
39	128.5	2.8	1051	2 G84316	ribonucleoside red
40	128.5	2.8	1238	2 A64596	hypothetical prote
41	128.5	2.8	1939	2 D97316	probable S-layer p
42	128	2.8	1250	2 F85862	hypothetical prote
43	128	2.8	1487	2 AG2560	hypothetical prote
44	127.5	2.8	936	2 S57637	hexon protein - hu
45	127.5	2.8	1067	2 D75625	probable extracell

ALIGNMENTS

RESULT 1

T42131
probable toxR-regulated lipoprotein tagA - Escherichia coli plasmid p0157
C:Species: Escherichia coli
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Nov-2000
C:Accession: T42131; T00210
R:Burland, V.; Shao, Y.; Perna, N.T.; Plunkett, G.; Sofia, H.J.; Blattner, F.R.
Nucleic Acids Res. 26, 4196-4204, 1998
A:Title: The complete DNA sequence and analysis of the large virulence plasmid of Esc
A:Reference number: 222068; PMID:98391744; PMID:9722840
A:Accession: T42131
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-898 <BUR>
A:Cross-references: EMBL:AF074613; PIDN:AAC70099.1
A:Experimental source: strain EDL933; serotype O157:H7
R:Makino, K.; Ishii, K.; Yasunaga, T.; Hattori, M.; Yokoyama, K.; Yatsudo, H.C.; Kubo
S.; Shinagawa, H.
DNA Res. 5, 1-9, 1998
A:Title: Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an enterohemor
A:Reference number: 214127; PMID:98290540; PMID:9628576
A:Accession: T00210
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 569-898 <MAX>
A:Cross-references: EMBL:AB011549; NID:d1204561; PIDN:BAA31757.1; PID:d1032718
A:Experimental source: strain EHEC O157:H7, substrain RMD 0509952
C:Genetics:
A:Genome: plasmid p0157
A:Note: L7031

Query Match
Best Local Similarity 100.0%; Score 4597; DB 2; Length 898;
Matches 863; Conservat 100.0%; Pred. No. 5.7e-292;
Mismatches 0; Indels 0; Gaps 0;

QY	1	ADNNSAIFYNTSQPINDLQSLAAEYKFAQSQILPAHPKEDSQPHLTSLRKSLLLRPV	60
DB	36	ADNNSAIFYNTSQPINDLQSLAAEYKFAQSQILPAHPKEDSQPHLTSLRKSLLLRPV	95
QY	61	KADDKTPQVVEARDNNKILGTLTLPSSSLPTDIYHLDGVPGGIDFTPHNGTKKIINT	120
DB	96	KADDKTPQVVEARDNNKILGTLTLPSSSLPTDIYHLDGVPGGIDFTPHNGTKKIINT	155
QY	121	VAEYKLSDSAGSGSIHSHLTNNALVEIHTANGRWDRDIYLPQGDPLEGKRVFVSAGYS	180
DB	156	VAEYKLSDSAGSGSIHSHLTNNALVEIHTANGRWDRDIYLPQGDPLEGKRVFVSAGYS	215
QY	181	STVYIGDKVTLVSGNTLLFKYNGQWFRSGELENRITYAHIWIAELPAHVIPLNL	240
DB	216	STVYIGDKVTLVSGNTLLFKYNGQWFRSGELENRITYAHIWIAELPAHVIPLNL	275
QY	241	VIKQNLGSLRNDIKIGAPGELLHTIDIGLITPDRDFDFAKDKEAHREYFQTIPIVSRM	300

|||||
276 VIRQNTLSGRINDIKIGAPGELLHTIDIGMLTTPDRDFADKDKAHEHYFOTIPVSRM 335
QY
301 IVNNYAPLHLKEVMLPTGELLTDMDPGNGWHSGTMRORIGKELVSHGIDNANYGLNSTA 360
Db
336 IVNNYAPLHLKEVMLPTGELLTDMDPGNGWHSGTMRORIGKELVSHGIDNANYGLNSTA 395
QY
361 GLGENSHPYVVAQAAHNSRGNTANGIQVHGSGGGGIVTLDSTLGNESHEVGHNYGLG 420
Db
396 GLGENSHPYVVAQAAHNSRGNTANGIQVHGSGGGGIVTLDSTLGNESHEVGHNYGLG 455
QY
421 HYVDFKGSVHRSANNNKSTWGDKKRIPNFPYQTNKESCLNNOQCEPDDGKHFGF 480
Db
456 HYVDFKGSVHRSANNNKSTWGDKKRIPNFPYQTNKESCLNNOQCEPDDGKHFGF 515
QY
481 DAMAGGSPFSAANRFTMYTPNSSAIIORFTENKAVFDSRSTGFSKWNADTOEMEPYHT 540
Db
516 DAMAGGSPFSAANRFTMYTPNSSAIIORFTENKAVFDSRSTGFSKWNADTOEMEPYHT 575
QY
541 IDRAEQITASVNELSEKMAELMAEYAVVYVHMNGNWTNRIYIPTASADNRGSILTIH 600
Db
576 IDRAEQITASVNELSEKMAELMAEYAVVYVHMNGNWTNRIYIPTASADNRGSILTIH 635
QY
601 EAGNSYLFINGDEKVSQYKSFVSDGOFWKEKROVVDTRARKPQFGVPVTVLVGY 660
Db
636 EAGNSYLFINGDEKVSQYKSFVSDGOFWKEKROVVDTRARKPQFGVPVTVLVGY 695
QY
661 DPEGLSSYIYPAMYGAYGFTYSDSONLSDNCOLOVDTKEGOLRPLRANHRANNTVMN 720
Db
696 DPEGLSSYIYPAMYGAYGFTYSDSONLSDNCOLOVDTKEGOLRPLRANHRANNTVMN 755
QY
721 KFHINVTESQTOATLVNCKNIIKDTSLTPAPEGLTYTVNGQALPAKENECCIVSVNSG 780
Db
756 KFHINVTESQTOATLVNCKNIIKDTSLTPAPEGLTYTVNGQALPAKENECCIVSVNSG 815
QY
781 KRYCLPVQORSGYSLPDWIVGOEYVDSGAKAKVLLSDWNLNSYNRIGFVGVNVPADMK 840
Db
816 KRYCLPVQORSGYSLPDWIVGOEYVDSGAKAKVLLSDWNLNSYNRIGFVGVNVPADMK 875
QY
841 KVKANNGOYLDPSKPRSMRVYK 863
Db
876 KVKANNGOYLDPSKPRSMRVYK 898
RESULT 2
T09438
toxR-regulated lipoprotein tagA - Escherichia coli plasmid p0157
C:Species: Escherichia coli
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 11-Jan-2002
C:Accession: T09438
R:Makino, K.; Ishii, K.; Yasunaga, T.; Hattori, M.; Yokoyama, K.; Yatsudo, H.C.; Kubota, S.; Shinagawa, H.
DNA Res. 5, 1-9, 1998
A:Title: Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an enterohemorrhagic E. coli O157:H7 strain
A:Reference number: Z14127; MUID:98290540; PMID:9628576
A:Accession: T09438
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-587 <MAK>
A:Cross-references: EMBL:AB011549; NID:d1204561; PIDN:BA031835.1
A:Experimental source: strain EHEC O157:H7, substrain RMD 0509952
C:Genetics:
A:Gene: tagA
A:Genome: plasmid p0157
Query Match 56.7%; Score 2604.5; DB 2; Length 587;
Best Local Similarity: 96.8%; Pred. No. 3.7e-162;
Matches 491; Conservative 2; Mismatches 7; Indels 7; Gaps 2;
QY 29 AQSQILPAHPKSGDSOPHLTSLRKSLLLVPRVPKADDKTPVQVEARDNNKILGTLTLYPP 88
Db 63 AASQILPAHPKSGDSOPHLTSLRKSLLLVPRVPKADDKTPVQVEARDNNKILGTLTLYPP 122

QY 89 SSLPDTIYHLDGVPBGIDFTPHNGTKKIINTVAEYVKNKLSDSAGSSSHSLTNALVEIH 148
Db 123 SSLPDTIYHLDGVPBGIDFTPHNGTKKIINTVAEYVKNKLSDSAGSSSHSLTNALVEIH 182
QY 149 TANGRWVDIYLPQPDLEGKRVFVSSAGYSSTVFYGDVKVTLVSGNTLLFKYVANGOWF 208
Db 183 TANGRWVDIYLPQPDLEGKRVFVSSAGYSSTVFYGDVKVTLVSGNTLLFKYVANGOWF 242
QY 209 RSGELNNRIYAQIHNSAELPAHVIIVPGLNLVTKQNLNLSGRINDIKIGAPGELLHTID 268
Db 243 RSGELNNRIYAQIHNSAELPAHVIIVPGLNLVTKQNLNLSGRINDIKIGAPGELLHTID 302
QY 269 IGMLTTPDRDFADKDKAHEHYFOTIPVSRMIVNNTAPLHLKEVMLPTGELLTDMDPGN 328
Db 303 IGMLTTPDRDFADKDKAHEHYFOTIPVSRMIVNNTAPLHLKEVMLPTGELLTDMDPGN 362
QY 329 GWHSGTMRORIGKELVSHGIDNANYGLNSTAGLGENSHPYVVAQAAHNSRGNYANGIQ 388
Db 363 GWHSGTMRORIGKELVSHGIDNANYGLNSTAGLGENSHPYVVAQAAHNSRGNYANGIQ 422
QY 389 VHGGSGGGGIVTLDSTLGNESHEVGHNYGLGHVVDGFKGSVHRSANNNKSTWGDGDKK 448
Db 423 VHGGSGGGGIVTLDSTLGNESHEVGHNYGLGHVVDGFKGSVHRSANNNKSTWGDGDKK 482
QY 449 RFIPIFYQTNKESCLNNOQCEPDDGKHFGFDAMAGSPFSAANRFTMYTPNSSAIQOR 508
Db 483 RFIPIFYQTNKESCLNNOQCEPDDGKHFGFDAMAGSPFSAANRFTMYTPNSSAIQOR 542
QY 509 FFENKAVF----DSRSTGFS---KWN 528
Db 543 FFENKAVSIAPPPASASGMIKKWN 569
RESULT 3
T09438
toxR-activated lipoprotein - Vibrio cholerae
N:Alternate names: toxR-activated gene A protein
C:Species: Vibrio cholerae
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 17-Nov-2000
C:Accession: T09438; A39108
R:Karolis, D.K.R.; Johnson, J.A.; Bailey, C.C.; Boedeker, E.C.; Kapur, J.B.; Reeves, Proc. Natl. Acad. Sci. U.S.A. 95, 3134-3139, 1998
A:Title: A Vibrio cholerae pathogenicity island associated with epidemic and pandemic A:Reference number: Z16672; MUID:98169509; PMID:9501228
A:Accession: T09438
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1002 <KAR>
A:Cross-references: EMBL:AF034434; NID:g3004923; PIDN:AAC12274.1; PID:g3004926
A:Experimental source: strain N16961
R:Parsot, C.; Taxman, E.; Mekalanos, J.J.
Proc. Natl. Acad. Sci. U.S.A. 88, 1641-1645, 1991
A:Title: ToxR regulates the production of lipoproteins and the expression of serum re A:Reference number: A39108; MUID:91156664; PMID:2000374
A:Accession: A39108
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-25 <PAR>
A:Cross-references: GB:M60658
C:Genetics:
A:Gene: tagA
A:Note: part of the pathogenicity island (VPI); associated with epidemic and pandemic C:Keywords: lipid binding; lipoprotein
Query Match 15.0%; Score 689.5; DB 2; Length 1002;
Best Local Similarity: 24.2%; Pred. No. 9.3e-37;
Matches 232; Conservative 151; Mismatches 276; Indels 295; Gaps 39;
QY 4 NSAITPNTSQINDLOGSLAAEVKFAQSQIIPAHBFEGDSOPHLTSLRKSLLLVPR-VKA 62
Db 137 NEIDYHTTEIEKKNYVSGLEGEVRFVQTHVI--SPEGRKNEPEITGRDALILFKPSIKN 194
QY 63 DDKTPVQVEARDNNKILGTLTLYPSSLPDIYHLDGVPBGIDFTPHNGTKKIINTVA 122

Db 105 SSSILMKIYSEDG---LTSKVYMKSPMLPKTDQPID-----ID----- 230
QY 123 EVNKLSDASGSIHSHLTHNNAIVEHTANGRWVRDYLPOGPDLECKMVRVFSAGYSST 182
Db 231 ----- 230
QY 183 VFYGDRTVLSVGNLTLFKYVNGWFRSGELENNR-ITYAHOIWSAELPAHVIPLVGLNLV 241
Db 231 ----- 259
QY 242 IKOGLNSGRL-----NDIKIGAPGELLHTIDIGMLTTPDRDFDAKDKAHR-EYFQTI 295
Db 260 FEDEN--GNLGIIESERIKFSAPSELLIONIDGLMYKPRGNIVIKELERTAVDYFOKY 317
QY 296 PVSRLMIVNNAAPLHLKEVMLPTGELLTDMDPGNGWHSCTHQRIGKLVSHGIDNANYG 355
Db 318 PYSKLIIFSDTPIHFEXITLPGVTYTEKSADIGWHQDMREAVKALVSTGINNANLG 377
QY 356 LNSTAGLGENSHPYVVAOLAHSRGNANGIOVHGSGGGGIVTLDSTLGNESHEVGH 415
Db 378 IVASSGYSQ-QYNRLTNHTAHTNIGYVNGVYVHGSGGGGIVTLENTLHNSHELGH 436
QY 416 NYGLGHYVDGFKGSVHRSANNNSTWGDGKKRFPNFPYSQNEKSCL--NNQCOEPF 473
Db 437 NYGLGHYVAG--GTSH---GPDTSWGDGYYKRFIANFDMKRSQSNIRPDQNEVYKPF 490
QY 474 -DGKFGFDAMAGGSPF--SAANRFTMTNPSAIIQRFENKAVFDSRSTGFSKNAD 530
Db 491 MDKYTYLWDMASGGYDHONGIISRYTLHHPYVARIQDMLKNGAVV---INNDRYWD-E 546
QY 531 TOEMEPYEHTIDRAEQITASVNELSESKMAELMAEYAVVYVHMNGNWTNRIIYPTASAD 590
Db 547 LKNIVYVYKCT-----NFKVPKIKGVPVVTILGVYDPDKIN-----PSQLYPPYTS-- 591
QY 591 NRGSIILTIN-----HEAGYN-----SYLFING-- 612
Db 592 NYGNIFDLEKPRSESSLKGMQYKQVYKQVNYLDRVNTHTMLVNRKEEIKRFSYLSPKGKK 651
QY 613 -----DEKVVSGYKKSFVSDGQFWKERDQVVDTRFARKPEQFGVPVTVLVG---Y 659
Db 652 FEFLGYEDIENKICTGGRSIHYLEDG---KKNPI-----ESKYNDF---LLSIDGGEIS 701
QY 660 YDPEGTL-----SSYTPAMYGA-----YGFYSDSDSONLS--- 690
Db 702 YVPDSTIGESKICSLKMSGTYYGAGFIKGNSCRQIDGVFMNGFQWAFVTLNGSGVNSTYTW 761
QY 691 DNDCOLOVDTEKGLR-FRLANHRANNTVMKPHINVPTEOPTQATLVCNNKILDTKSL 749
Db 762 SNECVLKIKDKDNNIESISIPNRIEKNQSNKIHLNRSERPIIDINVCGEHELTSIKV 821
QY 750 TPAPEGLTYTVNGQALPARENEGCIVSVNSGKRYCLPVGQSGYSLPDWIVGQEVYVDSG 809
Db 822 SDNPD-----IKLLKGPPII-----VQGEHGYT-----SYEPKLPFG 852
QY 810 AKAKVLLSDWNLN-YNRIGEFVGVNVPADMKKVAWNGQYL---DFS-KPRSHRVV 861
Db 853 -----WFKHYDNPEPKNEINHELKGRVND-----NDEYICRFNFSDSOREMKFV 897

RESULT 4
B82276
ToxR-activated gene A protein VC0820 [imported] - Vibrio cholerae (strain N16961 serogro
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: B82276
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P.
I., R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833; PMID:10952301
A:Accession: B82276

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1013 <HEI>
A:Cross-references: GB:AE004167; GB:AE003852; NID:g9555268; PIDN:AAF93983.1; GSPDB:G
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC0820
A:Map position: 1

Query Match 15.0%; Score 689.5; DB 2; Length 1013;
Best Local Similarity 24.2; Pred. No. 9.4e-37;
Matches 232; Conservative 151; Mismatches 279; Indels 295; Gaps 39;

QY 4 NSAIYNTSPINDLOGSLAAAEVKFAQSOLIPAHPEGDSOPHLTSLRKSLLLYRP-VYA 62
Db 148 NEDYHTTIERKNGVSEGEVRFVQTHVI--SPGRKNEPEIITGROALILFKPSIKN 205
QY 63 DDKTPVQVEARDNNKILGTLTLPSPSLPDTIYHLDGVPPEGIDFTPHNGTKKIINTVA 122
Db 206 SSSILMKIYSEDG---LTSKVYMKSPMLPKTDQPID-----ID----- 241
QY 123 EVNKLSDASGSIHSHLTHNNAIVEHTANGRWVRDYLPOGPDLECKMVRVFSAGYSST 182
Db 242 ----- 241
QY 183 VFYGDRTVLSVGNLTLFKYVNGWFRSGELENNR-ITYAHOIWSAELPAHVIPLVGLNLV 241
Db 242 -----ENKVVSYSNSYSAEIPWNKMKSGMSLH 270
QY 242 IKOGLNSGRL-----NDIKIGAPGELLHTIDIGMLTTPDRDFDAKDKAHR-EYFQTI 295
Db 271 FEDEN--GNLGIIESERIKFSAPSELLIONIDGLMYKPRGNIVIKELERTAVDYFOKY 328
QY 296 PVSRLMIVNNAAPLHLKEVMLPTGELLTDMDPGNGWHSCTHQRIGKLVSHGIDNANYG 355
Db 329 PYSKLIIFSDTPIHFEXITLPGVTYTEKSADIGWHQDMREAVKALVSTGINNANLG 388
QY 356 LNSTAGLGENSHPYVVAOLAHSRGNANGIOVHGSGGGGIVTLDSTLGNESHEVGH 415
Db 389 IVASSGYSQ-QYNRLTNHTAHTNIGYVNGVYVHGSGGGGIVTLENTLHNSHELGH 447
QY 416 NYGLGHYVDGFKGSVHRSANNNSTWGDGKKRFPNFPYSQNEKSCL--NNQCOEPF 473
Db 448 NYGLGHYVAG--GTSH---GPDTSWGDGYYKRFIANFDMKRSQSNIRPDQNEVYKPF 501
QY 474 -DGKFGFDAMAGGSPF--SAANRFTMTNPSAIIQRFENKAVFDSRSTGFSKNAD 530
Db 502 MDKYTYLWDMASGGYDHONGIISRYTLHHPYVARIQDMLKNGAVV---INNDRYWD-E 557
QY 531 TOEMEPYEHTIDRAEQITASVNELSESKMAELMAEYAVVYVHMNGNWTNRIIYPTASAD 590
Db 558 LKNIVYVYKCT-----NFKVPKIKGVPVVTILGVYDPDKIN-----PSQLYPPYTS-- 602
QY 591 NRGSIILTIN-----HEAGYN-----SYLFING-- 612
Db 603 NYGNIFDLEKPRSESSLKGMQYKQVYKQVNYLDRVNTHTMLVNRKEEIKRFSYLSPKGKK 662
QY 613 -----DEKVVSGYKKSFVSDGQFWKERDQVVDTRFARKPEQFGVPVTVLVG---Y 659
Db 663 FEFLGYEDIENKICTGGRSIHYLEDG---KKNPI-----ESKYNDF---LLSIDGGEIS 712
QY 660 YDPEGTL-----SSYTPAMYGA-----YGFYSDSDSONLS--- 690
Db 713 YVPDSTIGESKICSLKMSGTYYGAGFIKGNSCRQIDGVFMNGFQWAFVTLNGSGVNSTYTW 772
QY 691 DNDCOLOVDTEKGLR-FRLANHRANNTVMKPHINVPTEOPTQATLVCNNKILDTKSL 749
Db 773 SNECVLKIKDKDNNIESISIPNRIEKNQSNKIHLNRSERPIIDINVCGEHELTSIKV 832
QY 750 TPAPEGLTYTVNGQALPARENEGCIVSVNSGKRYCLPVGQSGYSLPDWIVGQEVYVDSG 809
Db 833 SDNPD-----IKLLKGPPII-----VQGEHGYT-----SYEPKLPFG 863

A:Cross-references: GB:AE001437; PIDN:AAK80209.1; PID:gl5025253; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC2252

```
Query Match          3.6%  Score 166.5; DB 2; Length 1217;
Best Local Similarity 18.6%  Pred. No. 0.018;
Matches 179; Conservative 121; Mismatches 319; Indels 343; Gaps 48;

QY 56 LVRPKA-----DDKTPQVARDNNKILGTLFLYPPSSLP-----DTIYHLDGVPEGI 106
   :||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 29 IIRPVSAKAPNKTISKITVKENARV-----SNLSAKLNGDTLIQVNLGDETDI 79

QY 107 -----DFPHNCTKLIINTVAEVNKLSDAGSSIHSLTNALV-----EIHTA 150
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 80 KICEPQVLKYDKRP-SGQSSDGLVDPNKIWN-TGNISSDNDSPMTITKMTIKIS 137

QY 151 NGRVVRDIYLPQGDLEGKAV---RFVSSAGYSTVFYGRKVTLSVGNLTLLFKYVNGOW 207
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 138 KSDLTMSVY-----DSGKQIVKQOSTASKSVSFTHNSGDR-----FYGINGYN 181

QY 208 FR-----SGELEN-NRITYAHIWSAELPAHWIVPGLMLVIKQNLGRLND----- 253
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 182 FKEDSNKGLMRNGTESYAGYQCHGSPFYNSDGYGLLVDSGGSFTIGDTSLOYSGIS 241

QY 254 -----IKGAPGELLHTIDI-----GMLTTP-----RORFDFAKDKKAHRE 290
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 242 KTDYDYLMLGSPKEVISESDVSGKAPFPKWATGFTNTQWNNNSLSGTGDNDEKLKS 301

QY 291 YFOT-----IPVSRMIV-----NNYAPLHLKEVMLP---TGELLTMDPGNGGWHIS 333
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 302 VLNTYRSKQIPIDNFCFLDFENKKGWODNYGEFKWNTDNFPDAONGQLKAYMD-SKGLKMT 360

QY 334 GTMRQRI-----CKELVSHG-----IDNANYGLNSTAGLGENSHPYV 370
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 361 GIMPRILADSEQARYVTSKQWMLPGDSASDYCSGKMMHNVFAISDV----- 409

QY 371 VAQLAAHNSRNYANGIOVHGSGGGIVTLDTSLGNEFSHEVGHNYGLGHY----- 422
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 410 -----RKWNWNIQ-----DAFDKGIVGFWDNEDCENYVFCNGMNMERA 450

QY 423 -VDGFKGSHRSAENNSTWG-----WDGDKRFPINFPYPSQTNK--- 462
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 451 IYDG-----QRAYKNQRWLSNRNYAGAQRYCYGHWMSGD-----ISTGDSMANORERM 500

QY 463 -SCLNNOCEPFDGKFGD--AMAGSP-----PSAARNFTMTYTPNSSAIIQRFENKAV 525
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 501 LSAVN-----LGEAKWMDTGGFNDGDPPTENYARWMEFSAFTP-----IFRVHGQDNRY 550

QY 516 -----FDSRSSTGFSKWNADQEWEPYEHTIDRAEOITA-----SVNELSEK 558
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 551 RYPWAFGSTAAEAAKAKMQLRYTLPIYIYSYDRSASOSGLGLVRLSMHMEYPNDSNAANDK 610

QY 559 MAELMAEYAVKVMNMGWNTNRNIYIP-----TASADNRGSILTINHEAGYNSY--- 607
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 611 EAWFGDIALVSPVNVQGTSEKSIYLPNGNIDITTGREYTGQ-OTINYAVDSINWSDIP 669

QY 608 LFING-----DEKVSOGYKKSFPVD-----CQFWKER 635
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 670 LFIKSGAIPTODFENYVGERKIIDVYDAFPDSDKATTFDYDDGTSYDYENGSYDQK 729

QY 636 DVVDTREARKPFQGVPTTLVGYVDEPTLSSYI-----YP 672
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 730 MTLOTSTDSKSVQFNIDKNT--GSTPD--LKDYIVKMHVKGNGAVTANGALTYQSYSD 785

QY 673 AMYAGYFTYSDSONLSDNDCQLQVDTKQGLRPLANHRANNTVMNKFHINTYPTESQP 732
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 786 ALKASAGEYASCT-----DIYGNVYIKVSSGDAKN-----INTVSCNPLP 826

QY 733 TOATLVCNKK-----ILDTKSLTPAPEGLT--YTVNGOALPAKENEGCIVSNKGRYCL 785
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 827 VTIITAAANPKGGTYGPGTQVSLTASKSDATIIYTL-----DGTPTVNSTK-YIA 875
```

```
QY 786 PV 787
Db 876 PI 877

RESULT 8
JC6027
115K outer membrane protein precursor - Bacteroides thetaiotaomicron
N:Alternate names: Susc protein
C:Species: Bacteroides thetaiotaomicron
C>Date: 31-Dec-1996 #sequence_revision 31-Dec-1996 #text_change 08-Oct-1999
C:Accession: JC6027
R:Reeves, A.R.; D'Elia, J.N.; Frias, J.; Salyers, A.A.
J Bacteriol. 178, 823-830, 1996
A:Title: A Bacteroides thetaiotaomicron outer membrane protein that is essential for
A:Reference number: JC6027; MUID:96146534; PMID:8550519
A:Accession: JC6027
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-1038 <REE>
A:Cross-references: GB:L49338; MID:gl100064; PIDN:AAA95938.1; PID:gl100065
A>Note: It is uncertain whether Met-1, Met-14 or Met-20 is the initiator
C:Comment: This protein is an integral outer membrane protein, and it is essential f
C:Genetics:
A:Gene: susc
C:Keywords: transmembrane protein
F:1-39/Domain: signal sequence #status predicted <SIG>
F:40-1038/Product: 115K outer membrane protein #status predicted <MAT>
F:1028-1038/Domain: transmembrane #status predicted <TM>

Query Match          3.5%  Score 161; DB 2; Length 1038;
Best Local Similarity 18.1%  Pred. No. 0.031;
Matches 180; Conservative 120; Mismatches 328; Indels 366; Gaps 46;

QY 16 NDLOGLAAEYKFAQSQILPAEPKGDSDPHLTSLRKSLLLVR-----PVKADDKTP---V 68
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 131 NDMTGSVMA-----INPDELSKG-----ITNAQDMLSGKIAGSVISNDGTGGGA 177

QY 69 QVEARDDNNKILGTLTYPPSSLPDIYHLDGVPGEIGDTPHNGTKKIIINTVAEVNKK-- 126
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 178 QIRIRGGSSL-----NASNDPLIVDGL---AID---NEGIKMANGLSWNPAD 221

QY 127 -----LSDASGSSIHSLTNALVEIHTANGRWVRDIYLPQGDLEGKMTYRFVSSAGYS 180
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 222 IETLVLDASATAYIGSRASNGVIITTKKGK-----NGQAPSYTNGSVSFS 270

QY 181 STVFYGRKRVTLVSGNTLLFKYVNGWFRSGELENNRITYAQHIWSAELPA-----HW 233
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 271 KT-----QKRYDVLSDGEYRA-----YANQLWGLKPLADLTANTDW 307

QY 234 -----IVPGLNLVIKOG---NLSGRNLNDIKIGAPGELLHTIDIGMLTTPRDRDFAK 283
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 308 QDQIFRTAYSTDHRYVSIINGGFKNLPYRS-----LGVTDDNGIVKTSNFR-RETA 356

QY 284 DREAUREYFQ---TIPVSRMIV---NNYAPLHLKE---VMLPTGELLTMDP---GNGG 330
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 357 SYNLAESPFDHHLKFNINAKFMNGKNRYADSCRWTRALAIDTPRPVYSNEDPTQFTGGY 416

QY 331 WHSGTMRQRIGKELVSHGIDNANYGLNSTAGLGE-----NSHPYVVAQLAAHNSRG 381
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 417 WQN-----INSTGSPNDWKYTSNPNSPQNPLAALKELKNDKG 454

QY 382 NYANGIOVHGSGGGGIVTLDTSLGNEFSHEVGHNYGL-----CHYVDGFKG 428
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 455 N-----SNDFFVGNVDYKPFHLPDLRLKHAISIGGEAEGTQT 491

QY 429 SVHRSAAENNSTGWGDG---KKRFIPNFPSPQTSNEKSLNNOCEPFDGKFGFDAMA 484
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 492 TIVSPVTFNNYTGNGDVTQYKYNLSYIYQYI--KSLGAND-----FDLMV 538

QY 485 GSPFSAANRFTMYTPNSSAIRFFENKAVPDSRSSTGFSK---WNADTQEMPEYHTI 541
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 539 GGEE-----QHFRN-----GFEEGGQGWSDYTOE---PHDA 566
```

```
QY 542 DRAEQITASVNELSEKMA---ELMAEYAVVYVHMNGNWTNRIYIPTASADNRGSILT 597
D 567 KUREOTAYATRNLTYSYFGRNLNYSILNRYLFTTRWDGS-----SRFSKDNRW----- 615
QY 598 INHEAGYNSYLFINGDEKVVSGYKKSFFVDSQGFQWKRERDVVDTRREARKPEQFGVPVTLV 657
D 616 -----GTFPSLALGWK---IKENFLKOVNLSDLKLR-----LWGCITGOO 654
QY 658 GYDPEGTLSYIYPMYGAIFY--TYSDDS---QNLSDNDOLQVDTKEGQLRFLANHR 713
D 655 NIGDDPAYLPLVYVNNYEAQYFPGDTYSTSRPKAFENLWKEKTTTWNAGLDFGLNGR 714
QY 714 ANNTVNMKPHINVTESQPTQATLVCCNKKILDTKSLTPAPEGLTY-----TVNGO 763
D 715 ITGGIDYF-----RKTWCYALRSPMILQCPDDEYRFTGKLYGFSINAK 763
QY 764 ALPAKE-----NEGCTIVNSG--KRYCLPVG----- 788
D 764 PIVTKDFTWDLSYNTWNNHETIKLTGGDDSDYVEAGDKISRGNTKVAQKVGYYAANS 823
QY 789 -----QSRGYSLPOMIYGOEYVYVDSGAKAKVLLSDNDLNSNRIETGERVGNV 835
D 824 FTVSRGNTKVAQKVGYYAANSFYIQQYIDENGKPI-----ENMFVDRNGN--GTID 874
QY 836 PADMKKYKAWNG-----QYLDFSKPRSMR 859
D 875 SGDKVYIYKPKAGDVLMLGLTSKMKQYKNFDFSFSLR 908

RESULT 9
S70172
toxin B - Clostridium difficile
C:Species: Clostridium difficile
C:Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 15-Oct-1999
C:Accession: S70172; S44271
R: von Eichel-Streiber, C.; Meyer zu Heringdorf, D.; Habermann, E.; Sartingen, S.
Mol. Microbiol. 17, 313-321, 1995
A:Title: Cloning in on the toxic domain through analysis of a variant Clostridium difficile
A:Reference number: S70172; MUID:96079281; PMID:749480
A:Accession: S70172
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-2367 <VON>
A:Cross-references: EMBL:223277; NID:g761713; PIDN:CAA80815.1; PID:g761714
A:Experimental source: isolate 1470
R: Sartingen, S.; von Eichel-Streiber, C.
submitted to the EMBL Data Library, July 1993
A:Description: Cloning and sequencing of an Clostridium difficile toxin B mutant.
A:Reference number: S44271
A:Accession: S44271
A:Molecule type: DNA
A:Residues: 1-1323,'N',1325-2367 <SAR>
A:Cross-references: EMBL:223277
C:Superfamily: cpl repeat homology
C:Keywords: cytotoxin

Query Match 3.5%; Score 159.5; DB 2; Length 2367;
Best Local Similarity 18.4%; Pred. No. 0.14;
Matches 185; Conservative 149; Mismatches 321; Indels 353; Gaps 49;

QY 4 NSAIYNTSOPINDLQGLAAEVKPAQSOILPAHPKEGDSQPHLSLRKSLLLVRPVKAD 63
D 1454 NSELQKNIPSYDSEK-----ENGFIENGSTKEG----LFVSELDDVYLISKVYMD 1501
QY 64 DKTP-----VQVEARDNNKILG-----TLTLYPPSSLPTIYKLDGV 101
D 1502 DSAPSGYGVYNNLKVYTKDNVNLITGYLKDDIKISLTLQDEKIKLNSVHLD-- 1559
QY 102 PEGGIDFTPHNGTKKIINTVAEYKNSLDASGSIHSLTNALVEIHTANGRWVRYDILP 161
D 1560 -ESG-----VAEILKPMNRKGS---TNTSDSLMSFLESNN---IKSIFV- 1596
```

```
QY 162 QGPDLEGKAVRFYSSAGY---SSTVFYG-----DRKVTLSVGNT 197
D 1597 ---NFQSNKIFILDANFIISGTTSIGQEFFICDENNNTOPIEIKFTNTLETNYTLVGNR 1653
QY 198 L-LFKYVNGWFRSGELENRRITYAQ-HWSAELPAHWIYVGLNLVLIKQNLSSG----- 249
D 1654 QNMIVENPYDLDDSGDISSTVINFSQKLYGIDSCVKNVVISNIYTDINITYETNN 1713
QY 250 -----RLNDIKI-----CAPGELLHTIDIGMLTTPDRDP--DFAK 283
D 1714 TYPEVIVLDANYINEKINVININDLSIRYMSNDGNDFILMSTSEENKVSQVIRFVNFVK 1773
QY 284 DKE-AHREYF-----QTPVSRMIVANNYAPLHLKEVAL----- 315
D 1774 DKTLANKLSNFSKDKDVPVSEIL-SFTPSYEDGLIGTDLGLVSLYNEKFYNFNGHM 1832
QY 316 -----PTGELLTD-----MDPGNGGHSHTMRQIRIGKELYVSHGI 349
D 1833 VSGLIYINDSLYFKPPVNNLITGFTVGDGKYFNPINGGAAS-----IGETI---I 1882
QY 350 DNANYGLNSTAGLGENSHPVVVAQLAAHNSRGNY-----ANGIQVHGSGGGGIVTLDS 404
D 1883 DDKNYFNQSG-----VLQGVFSTEDGFKYFAPAN-----TIDEN 1918
QY 405 LGNE---FSHEVGHNYGLCHYVDGFKGSVHRSANNSTWGW---DQDKRRFIPNFPY 459
D 1919 LEGEADFTKLIIDENIYFEDNYRGAVE-----WKELDGMHYFSPE----- 1962
QY 459 TNEKSLNNOCQPFDFGKFKGFDMAGGSPFSAANRF--TWYTPNSAIIOQFF---EN 512
D 1963 -----TGKAFKGLNIGDDKYYFNSDGVNQKGFVSINDN 1996
QY 513 KAVFDSRS--STGFSK-----WNADTOEMEPYEHTIDRAEOITASVNELSEKMAELMA 564
D 1997 KHYFDDSGVMKVGYTEIDGKHFFAENGQIGVFNTEGFKYFAHNEDLGNESSEIS 2056
QY 565 EYAVVYVHMNGNWTNRIYIPTASADNRGSIILINHEAGYNSYLFINGDEKVVSGYKKS 624
D 2057 YSGIL-----NFNKKIYY---FDSFTAVVQKLEDDGSKYFF---DEDTAEVIGLS 2103
QY 625 FVSDGFQWKRERDVVDTRREARKPEQFG-VPVTVLVGYVDPEGLTSSYIYPAMYGAFTYS 683
D 2104 LINDGQIYFNDGDI-----MOVGFVTINDKVFYFSDSGIIES-----GVONI 2145
QY 684 DDSQNLSDNDQLOV---DTRKQGLRFLANHRANNTVMNKHINVTESQPTQATLVCN 740
D 2146 DNYFYIDDNGIYQIGVYFSDGKYFAPA-----NTVNDNIY----- 2183
QY 741 NKILDTKSLTPAPE-----GLTYVNGQALPAKENEGCIYVNSGKRYCLPVGQRS--GY 793
D 2184 GQAVEYSGLVVRGVEDVYFGETYTIETGIWYDMENE-----SDRYFYFETKKACKGI 2236
QY 794 SLPDWIVGQEVYVDSGAKAKVLLSDNDNLSY---NRIGEF-VGNVNPAD 838
D 2237 NLIDDI---KYFDEKGIIMRTGLISFNNYFNENGEIOFGYINIED 2281

RESULT 10
F97255
fusion of alpha-glucosidase (family 31 glycosyl hydrolase) and glycosidase (Trea/Mals
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: F97255
R: Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: F97255
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1157 <KUR>
A:Cross-references: GB:AE001437; PIDN:AAK80833.1; PID:g15025938; GSPDB:GN00168
```


A; Experimental source: Clostridium acetobutylicum ATCC84
C; Genetics:
A; Gene: CAC2891

Query Match	3.48	Score 156.5	DB 2	Length 1157
Best Local Similarity	19.3%	Pred. No. 0.073		
Matches 198	Conservative 120	Mismatches 335	Indels 371	Gaps 56

Qy	60	VKADDTKPVQVEARDDNKK-----ILGTLTILYPPSSLPTDIYHLHGVPEGGIDFTPHNGTK	115
Db	35	VKADTNVNVNKNSSSSOKKFHAKLANGTILKIKGKDETIIRICEPQVFKYDKP-NGKS	93
Qy	116	KIINTAEVKNKLSDAGSSITHSLTNNALVEIHTANGRWVRDI-----YLPQGPDI	166
Db	94	SKOTLVVDPNK-KWSTGNIVSSOLKSPMW---ITTKKWLKINKEDELSILVY-----DL	144
Qy	167	EGRKVRFSVAGSYSSYVF---YGRDKYVTLVSGNTLLEFKYNGWFR---SGELEN-NRI	218
Db	145	QKGLLLKQDSTASKTASFTTHNSGDR-----FYGINGYFNQEDSSKGLMRNCTES	193
Qy	219	TYAQHWSAELPAHWIYVGLNVLVKQNLGSLRLND-----IKTGAGEL	262
Db	194	YIAGYQHGCGSPFVMSNDGYGLLVDDSGGSFTIGDTSLKYDGIKSTDTDIYVMVNGKEI	253
Qy	263	LLHTIDIG-----MLTTPDRDF-----DFAKDKEAHREY-FQTIPIVSRMI	303
Db	254	LSEEDSVGRAPMPFKKANGFTNTQGWDNLSCTGNDKALKSVINTYRSKQLPIDNFC	313
Qy	302	V-----NNYAPLHLKEVMLP---TGELLTDMDPGNGWHSGTMRQRI-----GKE	343
Db	314	LDFDWWKKGQDNYGEFKWNTDNFPDSONGQLKAYMD-SKGLKMTGINKPRILADSKQGRY	372
Qy	344	LVSHG-----IDNANYGLNSTAGLGENSEHPYVVAQLAAHNSRGNTANG	386
Db	373	VTSKGWMLPGDSEASDYCSGKQMDENVPAL-----PQV-----RWMMNN	412
Qy	387	IQVHGSGGGGIVTLDDSTLGNFESHEVGHNYGLGHY-----VDGFKGSVHRSAAEN	437
Db	413	IQ---GAFDKGIYGF-----WQDECDENVFONFGNNMERAIYDG-----QRRHKN	456
Qy	438	NSTWG-----WDGKKRFIPNTPYPSQTNEK-----SCLNNOCQEPFPGHK	477
Db	457	QRVMSLRNRYAGAQRYSGHWSGD-----ISTGFDSPANQRMLSAVN-----LGEAK	506
Qy	478	FGED--AMAGSP---FSAANRRMTKTPNSSALIQRFF--ENKA---VFDSSSSIGFS	525
Db	507	WGHDTGFGNGDPTPEYIARMEFSATFP-----IFRVHGQDNKYVPWAFGSTAEATAK	561
Qy	526	KWNADTQEMPEYHTIDRAEQITA-----SYNELSFSKMAELMAEYAVVKVHM	573
Db	562	KAMQLRVTLPIYISYDRSASQGLGLVRLSMVEVPNDSNAANDKEAMWFGDYLVPV	621
Qy	574	WNGWNTRNIIYP-----TASADNRSGILTIINHEAGYSY---LFLNG-----	612
Db	622	QEGQTSKIYLPNGWNIDYTTGREYTG-QTINYAVDSTNNWSDIPLFKISGAIITDQFE	680
Qy	613	---DEKVVYQCYKKSF-----VSDGQFWKRDVVDTEARKPEQFG	650
Db	681	NYVGEKIIDYVDAPFGNEASSFDYDDCGTSYNYENGSTFDQKNTLERAKDLKSVQFN	740
Qy	651	VPVTTLVGYVDPEGTLSYIYP-----AMYGAYGFTYSDDSONLSNDXC	694
Db	741	ISPKT--GY--KSDLKNYIVKMHVKSSGDVTVGRRITRYASY-----DELKNAQEGY	791
Qy	695	QLOVDTEKQLRFL-ANIRANNIVNKKFHNVPTESOPTQATLVCNNKIL-----DTKSL	749
Db	792	VVGDTDTYGSVYIIRKVSACHDKN-----INVP-----CNOVOLTAYADVK--	830
Qy	750	TPAPEGLTYVNGALPAKEN-----EGCIVSVNSCKRYCLPVGORSGLSDPMI	799
Db	831	-----GGTYT--SPQKVSUKASDPNNAIYYTLDGTAPTIVNSTK-----YTGp	870
Qy	800	VGQEVYVDSGAKAKVLLSDWNLNSYNRIGEEFVGNNVPADMKVKVAINNGOYLDGSKPSMR	859

```

      871      : :| |      | : : |      : | | :      : :
      ----ITIDSKTLKIVRD-----ANGNESDV-----FTPEQYTY-----IK 903

Db      871      : :| |      | : : |      : | | :      : :
      860      VYVK 863      |      |      |
      904      VHYK 907      |      |      |

RESULT 11
S54500
alpha,alpha-trehalase (EC 3.2.1.28) - yeast (Saccharomyces cerevisiae)
N:Alternate names: acid trehalase ATHI; protein YP9367.06; protein YPR026w
C:Species: Saccharomyces cerevisiae
C>Date: 08-Jul-1995 #sequence_revision 19-Oct-1995 #text_change 29-Oct-1999
C:Accession: S54500; S60459; S70514
R:Badcock, K.; Churcher, C.M.
submitted to the EMBL Data Library, May 1995
A:Reference number: S54059

```

RESULT 11

alpha, alpha-trehalase (EC 3.2.1.28) - yeast (*Saccharomyces cerevisiae*)
N:Alternate names: acid trehalase ATH1; protein YP9367.06; protein YPR026w
C:Species: *Saccharomyces cerevisiae*
C:Date: 08-Jul-1995 #sequence.revision 19-Oct-1995 #text_change 29-Oct-1999
C:Accession: S54500; S60459; S70514
R:Badcock, K.; Churcher, C.M.
submitted to the EMBL Data Library, May 1995
A:Reference number: S54059
A:Accession: S54500
A:Molecule type: DNA
A:Residues: 1-1211 <BAD>
A:Cross-references: EMBL:Z49274; MID:g809585; PIDN:CAA89280.1; PID:g809591; MIPS:YPR026w
A:Experimental source: strain AB972
R:Destruelle, M.; Holzer, H.; Klionsky, D.J.
Yeast 11, 1015-1025, 1995
A:Title: Isolation and characterization of a novel yeast gene, ATH1, that is required for growth of *Saccharomyces cerevisiae* on trehalose
A:Reference number: S60459; MUID:96076626; PMID:7502577
A:Accession: S60459
A:Molecule type: DNA
A:Residues: 131-1167, 'FAG', 1171 <DES>
A:Cross-references: EMBL:X84156; MID:g1061283; PIDN:CAA58961.1; PID:g1061284
A:Note: the authors translated the codon GTT for residue 18 as Leu, GAG for residue : 19 as Leu, GAG for residue : 20 as Leu, GAG for residue : 21 as Leu, GAG for residue : 22 as Leu, GAG for residue : 23 as Leu, GAG for residue : 24 as Leu, GAG for residue : 25 as Leu, GAG for residue : 26 as Leu, GAG for residue : 27 as Leu, GAG for residue : 28 as Leu, GAG for residue : 29 as Leu, GAG for residue : 30 as Leu, GAG for residue : 31 as Leu, GAG for residue : 32 as Leu, GAG for residue : 33 as Leu, GAG for residue : 34 as Leu, GAG for residue : 35 as Leu, GAG for residue : 36 as Leu, GAG for residue : 37 as Leu, GAG for residue : 38 as Leu, GAG for residue : 39 as Leu, GAG for residue : 40 as Leu, GAG for residue : 41 as Leu, GAG for residue : 42 as Leu, GAG for residue : 43 as Leu, GAG for residue : 44 as Leu, GAG for residue : 45 as Leu, GAG for residue : 46 as Leu, GAG for residue : 47 as Leu, GAG for residue : 48 as Leu, GAG for residue : 49 as Leu, GAG for residue : 50 as Leu, GAG for residue : 51 as Leu, GAG for residue : 52 as Leu, GAG for residue : 53 as Leu, GAG for residue : 54 as Leu, GAG for residue : 55 as Leu, GAG for residue : 56 as Leu, GAG for residue : 57 as Leu, GAG for residue : 58 as Leu, GAG for residue : 59 as Leu, GAG for residue : 60 as Leu, GAG for residue : 61 as Leu, GAG for residue : 62 as Leu, GAG for residue : 63 as Leu, GAG for residue : 64 as Leu, GAG for residue : 65 as Leu, GAG for residue : 66 as Leu, GAG for residue : 67 as Leu, GAG for residue : 68 as Leu, GAG for residue : 69 as Leu, GAG for residue : 70 as Leu, GAG for residue : 71 as Leu, GAG for residue : 72 as Leu, GAG for residue : 73 as Leu, GAG for residue : 74 as Leu, GAG for residue : 75 as Leu, GAG for residue : 76 as Leu, GAG for residue : 77 as Leu, GAG for residue : 78 as Leu, GAG for residue : 79 as Leu, GAG for residue : 80 as Leu, GAG for residue : 81 as Leu, GAG for residue : 82 as Leu, GAG for residue : 83 as Leu, GAG for residue : 84 as Leu, GAG for residue : 85 as Leu, GAG for residue : 86 as Leu, GAG for residue : 87 as Leu, GAG for residue : 88 as Leu, GAG for residue : 89 as Leu, GAG for residue : 90 as Leu, GAG for residue : 91 as Leu, GAG for residue : 92 as Leu, GAG for residue : 93 as Leu, GAG for residue : 94 as Leu, GAG for residue : 95 as Leu, GAG for residue : 96 as Leu, GAG for residue : 97 as Leu, GAG for residue : 98 as Leu, GAG for residue : 99 as Leu, GAG for residue : 100 as Leu, GAG for residue : 101 as Leu, GAG for residue : 102 as Leu, GAG for residue : 103 as Leu, GAG for residue : 104 as Leu, GAG for residue : 105 as Leu, GAG for residue : 106 as Leu, GAG for residue : 107 as Leu, GAG for residue : 108 as Leu, GAG for residue : 109 as Leu, GAG for residue : 110 as Leu, GAG for residue : 111 as Leu, GAG for residue : 112 as Leu, GAG for residue : 113 as Leu, GAG for residue : 114 as Leu, GAG for residue : 115 as Leu, GAG for residue : 116 as Leu, GAG for residue : 117 as Leu, GAG for residue : 118 as Leu, GAG for residue : 119 as Leu, GAG for residue : 120 as Leu, GAG for residue : 121 as Leu, GAG for residue : 122 as Leu, GAG for residue : 123 as Leu, GAG for residue : 124 as Leu, GAG for residue : 125 as Leu, GAG for residue : 126 as Leu, GAG for residue : 127 as Leu, GAG for residue : 128 as Leu, GAG for residue : 129 as Leu, GAG for residue : 130 as Leu, GAG for residue : 131 as Leu, GAG for residue : 132 as Leu, GAG for residue : 133 as Leu, GAG for residue : 134 as Leu, GAG for residue : 135 as Leu, GAG for residue : 136 as Leu, GAG for residue : 137 as Leu, GAG for residue : 138 as Leu, GAG for residue : 139 as Leu, GAG for residue : 140 as Leu, GAG for residue : 141 as Leu, GAG for residue : 142 as Leu, GAG for residue : 143 as Leu, GAG for residue : 144 as Leu, GAG for residue : 145 as Leu, GAG for residue : 146 as Leu, GAG for residue : 147 as Leu, GAG for residue : 148 as Leu, GAG for residue : 149 as Leu, GAG for residue : 150 as Leu, GAG for residue : 151 as Leu, GAG for residue : 152 as Leu, GAG for residue : 153 as Leu, GAG for residue : 154 as Leu, GAG for residue : 155 as Leu, GAG for residue : 156 as Leu, GAG for residue : 157 as Leu, GAG for residue : 158 as Leu, GAG for residue : 159 as Leu, GAG for residue : 160 as Leu, GAG for residue : 161 as Leu, GAG for residue : 162 as Leu, GAG for residue : 163 as Leu, GAG for residue : 164 as Leu, GAG for residue : 165 as Leu, GAG for residue : 166 as Leu, GAG for residue : 167 as Leu, GAG for residue : 168 as Leu, GAG for residue : 169 as Leu, GAG for residue : 170 as Leu, GAG for residue : 171 as Leu, GAG for residue : 172 as Leu, GAG for residue : 173 as Leu, GAG for residue : 174 as Leu, GAG for residue : 175 as Leu, GAG for residue : 176 as Leu, GAG for residue : 177 as Leu, GAG for residue : 178 as Leu, GAG for residue : 179 as Leu, GAG for residue : 180 as Leu, GAG for residue : 181 as Leu, GAG for residue : 182 as Leu, GAG for residue : 183 as Leu, GAG for residue : 184 as Leu, GAG for residue : 185 as Leu, GAG for residue : 186 as Leu, GAG for residue : 187 as Leu, GAG for residue : 188 as Leu, GAG for residue : 189 as Leu, GAG for residue : 190 as Leu, GAG for residue : 191 as Leu, GAG for residue : 192 as Leu, GAG for residue : 193 as Leu, GAG for residue : 194 as Leu, GAG for residue : 195 as Leu, GAG for residue : 196 as Leu, GAG for residue : 197 as Leu, GAG for residue : 198 as Leu, GAG for residue : 199 as Leu, GAG for residue : 200 as Leu, GAG for residue : 201 as Leu, GAG for residue : 202 as Leu, GAG for residue : 203 as Leu, GAG for residue : 204 as Leu, GAG for residue : 205 as Leu, GAG for residue : 206 as Leu, GAG for residue : 207 as Leu, GAG for residue : 208 as Leu, GAG for residue : 209 as Leu, GAG for residue : 210 as Leu, GAG for residue : 211 as Leu, GAG for residue : 212 as Leu, GAG for residue : 213 as Leu, GAG for residue : 214 as Leu, GAG for residue : 215 as Leu, GAG for residue : 216 as Leu, GAG for residue : 217 as Leu, GAG for residue : 218 as Leu, GAG for residue : 219 as Leu, GAG for residue : 220 as Leu, GAG for residue : 221 as Leu, GAG for residue : 222 as Leu, GAG for residue : 223 as Leu, GAG for residue : 224 as Leu, GAG for residue : 225 as Leu, GAG for residue : 226 as Leu, GAG for residue : 227 as Leu, GAG for residue : 228 as Leu, GAG for residue : 229 as Leu, GAG for residue : 230 as Leu, GAG for residue : 231 as Leu, GAG for residue : 232 as Leu, GAG for residue : 233 as Leu, GAG for residue : 234 as Leu, GAG for residue : 235 as Leu, GAG for residue : 236 as Leu, GAG for residue : 237 as Leu, GAG for residue : 238 as Leu, GAG for residue : 239 as Leu, GAG for residue : 240 as Leu, GAG for residue : 241 as Leu, GAG for residue : 242 as Leu, GAG for residue : 243 as Leu, GAG for residue : 244 as Leu, GAG for residue : 245 as Leu, GAG for residue : 246 as Leu, GAG for residue : 247 as Leu, GAG for residue : 248 as Leu, GAG for residue : 249 as Leu, GAG for residue : 250 as Leu, GAG for residue : 251 as Leu, GAG for residue : 252 as Leu, GAG for residue : 253 as Leu, GAG for residue : 254 as Leu, GAG for residue : 255 as Leu, GAG for residue : 256 as Leu, GAG for residue : 257 as Leu, GAG for residue : 258 as Leu, GAG for residue : 259 as Leu, GAG for residue : 260 as Leu, GAG for residue : 261 as Leu, GAG for residue : 262 as Leu, GAG for residue : 263 as Leu, GAG for residue : 264 as Leu, GAG for residue : 265 as Leu, GAG for residue : 266 as Leu, GAG for residue : 267 as Leu, GAG for residue : 268 as Leu, GAG for residue : 269 as Leu, GAG for residue : 270 as Leu, GAG for residue : 271 as Leu, GAG for residue : 272 as Leu, GAG for residue : 273 as Leu, GAG for residue : 274 as Leu, GAG for residue : 275 as Leu, GAG for residue : 276 as Leu, GAG for residue : 277 as Leu, GAG for residue : 278 as Leu, GAG for residue : 279 as Leu, GAG for residue : 280 as Leu, GAG for residue : 281 as Leu, GAG for residue : 282 as Leu, GAG for residue : 283 as Leu, GAG for residue : 284 as Leu, GAG for residue : 285 as Leu, GAG for residue : 286 as Leu, GAG for residue : 287 as Leu, GAG for residue : 288 as Leu, GAG for residue : 289 as Leu, GAG for residue : 290 as Leu, GAG for residue : 291 as Leu, GAG for residue : 292 as Leu, GAG for residue : 293 as Leu, GAG for residue : 294 as Leu, GAG for residue : 29

Query Match 3.4%; Score 154; DB 2; Length 1211;
Best Local Similarity 18.4%; Pred. No. 0.11;
Matches 143; Conservative 104; Mismatches 265; Indels 266; Gaps 35

[illegible]

Db 392 NVLTEDOPKIIVKYGINTEFNKEQO--DNTNIGL-----AKMIA 434
QY 377 HNSRGNYANGIQVH-----
Db 435 LNSKNGYEKLLSSHKRAWYDLYNDAFTEIPSDSLEMTARSSLFHLLANTRDYNVSSDRG 494
QY 391 ---GGSG-----GGGIYVLDSTLGN-----FSHEVGHNYGLGHYVDFKGSVHRSSENN 437
Db 495 LPVGVSGLSDSYSGOMFWDADHWEFALLPFFPNVAQN-----MNNYRNATHSOAKLN 548
QY 438 NSTGWDCDKRFRTPNFYPSPTNKSCLNNOCQEPFDCGHKFGFDAMAGGS-----487
Db 549 AEKYGPCALYPTWSTGKYANCTSTGPCVDYEHINVDVANAFSIIYLNHGEIDDEYLRY 608
QY 488 ---PF--SAARFMYTPT--NSAIIQRFPE-----NKAVFDSRSTGSKKNAD 530
Db 609 TWPPIKNAAQFFATYVKNYNSLSLYETYNLTDPDEFANHINNGAFTNAGIKTLIKWATD 668
QY 531 TOEMEPYHTIDRAEQITASVNSLSESKMAELMAEYAVVYHMMNGNWTNRIYIPTASAD 590
Db 669 -----TONHGEVVDPAWSEI-----SKDIYIPRSSN 696
QY 591 NRGSILTINHEAGVNSYL-----PINGDEKVSQGYKKSFSVDSQGFWR 635
Db 697 -----ITLEY-SGMNSSVEIKOADVTLMVYPLGVIN-DESLNNAIK-----DLYYYSER 744
QY 636 DWDVTRARKPEQ-FGVPTTLVGYDPEGLTSSYIYPAMYGAYGFTYSDDSQNLSDN 692
Db 745 -----QSASGPANTYPPVFAAAGLLNHGSSSQSLYKSVLPYLRAPFAQFSESDDN 797

RESULT 12
S10317
toxin B - Clostridium difficile
C:Species: Clostridium difficile
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 15-Oct-1999
C:Accession: S10317; S21894; S22434
R:Barroso, L.A.; Wang, S.Z.; Phelps, C.J.; Johnson, J.L.; Wilkins, T.D.
Nucleic Acids Res. 18, 4004, 1990
A:Title: Nucleotide sequence of Clostridium difficile toxin B gene.
A:Reference number: S10317; MUID:90326540; PMID:2374729
A:Accession: S10317
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-2366 <BAR>
A:Cross-references: EMBL:X53138; NID:q40442; PIDN:CAA37298.1; PID:q40443
R:Eichel-Streiber, C.; Laufenberg-Feldmann, R.; Sartingen, S.; Schulze, J.; Sauerborn, M.
submitted to the EMBL Data Library, July 1991
A:Description: Comparative analysis of Clostridium difficile toxins A and B.
A:Reference number: S21894
A:Accession: S21894
A:Molecule type: DNA
A:Residues: 1271-2366 <EIC>
A:Cross-references: EMBL:X60984; NID:q40445; PIDN:CAA43299.1; PID:q40446
R:von Eichel-Streiber, C.; Laufenberg-Feldmann, R.; Sartingen, S.; Schulze, J.; Sauerborn, M.
Mol. Gen. Genet. 233, 260-268, 1992
A:Title: Comparative sequence analysis of the Clostridium difficile toxins A and B.
A:Reference number: S22434; MUID:92293124; PMID:1603068
A:Accession: S22434
A:Molecule type: DNA
A:Residues: 1791-2366 <VON>
A:Cross-references: EMBL:X60984
C:Genetics:
A:Gene: toxB
C:Superfamily: cpl repeat homology
C:Keywords: cytotoxin

Query Match 3.3%; Score 153.5; DB 2; Length 2366;
Best Local Similarity 18.6%; Pred. No. 0.35;
Matches 186; Conservative 155; Mismatches 321; Indels 339; Gaps 51;
QY 4 NSAIYPTNSQIPNDLQSLAAEVKFAQSQILPAHPKEGDSQPHLTSLRKSLLLVPRVKAD 63

Db 1453 NSELAKNIPYSVDSEK-----ENGFINSGTKEG-----LFVSELPDVLISKVYMD 1500
QY 64 OKTP-----VQVEARDNRKILG-----TTLTYPPSSLPDIYHLD-- 99
Db 1501 DSKPSFGYSSNNLKDVKVIKDNVNILTYGLKDDIKISLTLQDEKTIKNSVHDES 1560
QY 100 GYPEGGIDTPHNGTKIINTVAEVNKLSDAGSGSIHSLTNALYEHHTANGRVRD-- 157
Db 1561 GVAE-----ILKFMNRKGTN-TSDSLMSFLESNFKSIFFVNFQSNIFILDAN 1609
QY 158 -----YVLPQPDLEGRKAVRFVSSAGYSTVFYGDRTVL-----S 193
Db 1610 FIISGTTSIGOFEFICDENDNIQYFIK-NTLENTLYTVGNRMIVENYDLDSDG 1668
QY 194 VGNLL-----FKYVQWFRSGELEN--NRITYAHINSAEL---PAH-----WIVPG 237
Db 1669 ISSTVINFQKLYG-----IDSCVKNVISPNIYDEINITPVYETNTTYPEVIVLD 1721
QY 238 LNLVIKOGNLSGLNDIKI-----GAPGELLHIDIGMLTTPDRDF-DFAKDE-AHR 289
Db 1722 ANYINEKINV--INDLSRYVWSNDGDNDFILMSTSEENKVSQVKIRFVNVFKDKTLANK 1779
QY 290 EYF-----QTIPIVRMIVNNYAPLHLKEVNL-----315
Db 1780 LSNFSDKQDVPVSEIL-SFTPSYVEDGLIGDGLVSLYNEKFYINNFGMWSGLIYI 1838
QY 316 -----PTGELLTD-----MDPGNGWHSGTMRQIRIGKELVSHGIDNANYGL 356
Db 1839 NDSLYFKPPVNNLITGFTVGDQKYPFNPINGGAAS-----IGETI-----IDDKNTYF 1888
QY 357 NSTAGLGESHPTVYVAQLAAHNSRNY-----ANGIQVHGGSGGGIVTLDSTLGN-- 408
Db 1889 MQSG-----VLQGTGVFSTEDGFKYFAPAN-----TLDNLEGEAID 1924
QY 409 FSHEVGHNYGLHYDGFKGSVHRSSENNNSWGW---DGDKKRFIPNFPYPSQTEKSL 465
Db 1925 FTOKLIIDENIYDFDNYRGAVE-----WKELDGEHMFYSPE-----1961
QY 466 NNOCQEPFDGKFGFDAMAGGSPFSAANRF--TWYTPNSSAIIQRF---ENKAVFDSR 519
Db 1962 -----TGRAFKGLNIGDYKYVFNDSGVNMQKGFVSINDKNKHFFDS 2002
QY 520 S--STGFSK-----WNADTQEMEPYEHYIDRAEQITASVNSLSESKMAELMAEYAVVKV 571
Db 2003 GVMKVGYTEIDGKHFFAENGEMQGVNFTEDGFKYFAHNEDLGNKEGEETISYSGIL-- 2060
QY 572 HMNGNWTNRIYIPTASADNRGSILTINHEAGVNSYLFINGDEKVVSGYKKSFEVSDGOF 631
Db 2061 -----HFNKKIYY--FDSFTRAVGWKLEDDGSKYFF--DEDTAEAYIGLSLNDGOY 2109
QY 632 WKERDVVDTRARKPEQFG-VPVTTLVGYDPEGLTSSYIYPAMYGAYGFTYSDDSQNL 690
Db 2110 YFNDDGI-----MQVGFVINDKVFYFSDSGIIES-----GVQINDDNYFYI 2151
QY 691 DNDQCOLV---DTKEQLRFLRANRANTYNNKFNINVTESQPTQATLVCCNKILDTK 747
Db 2152 DDNGIYQIGVIGVTSDDGKYKYPAPA-----NTVNDNIY-----GQAVEYS 2189
QY 748 SLTPAPE-----GLTYTVNGQALPAKENEGCIVSVNSGKRYCLPVGORS--GYSIPDWTIV 800
Db 2190 GLYRVGEDVYFGETYTIETGWIYDENE-----SDKYYFNPEPKKACKGINLIDDI- 2241
QY 801 GQEVYVDSGAKAVLLSDMDNLSY--NRIGEF-VGNVNPAD 838
Db 2242 --KYTFDEKIGMTGLISFENNYYFENGENGMQFGYINIED 2280

RESULT 13
A83412
hypothetical protein PA1874 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: A83412
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Berman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, J.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: A83412
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2468 <STO>
A:Cross-references: GB:AE004613; GB:AE004091; NID:99947856; PIDN:AAG05263.1; GSPDB:GN001
A:Experimental source: strain PAO1
C:Genetics:
A:Gene: PA1874

Query Match 3.2%; Score 148.5; DB 2; Length 2468;
Best Local Similarity 19.5%; Pred. No. 0.79;
Matches 182; Conservative 90; Mismatches 301; Indels 361; Gaps 45;

QY 10 NTSQPINLQSLAEVFAQSILP-----AHP-----KEGDSQPH----- 46
DB 324 NNSTPVTVEADPTAPADPTQVAPDQVAPDQVSSVTGNAEPGATVGVDTDGDQDPTVVVGGP 383
QY 47 -----LTSRLKSLVRYKADKTPQVEARD-----DNNKIL----- 80
DB 384 GSFEVPLNPPLNGETVTVITDPAGNSSTPTAEAPDPAPQVNASGSLVSGTAEAG 443
QY 81 -----GTLTLYPPSLPD-----TIYHLDGV 101
DB 444 VTIVITDGNPNIGOTSADANGNSFTPGSQLPDGVVNVVVARDAAGNSPATITVDGV 503
QY 102 PEGGIDFTPHNGTKKIINTVAEVNK---LSASGSSIHSHLTNNALVEIHTANGRWVRI 158
DB 504 APNAVPEFSGSE--LSGTAEPSGVTLTGNGHPICQTAD-----ANGW---S 550
QY 159 YLPOGPDLEGKMYRVS--SAGYSS-----TVFYGDRKVTLSVGNWTLFFKYVNGOMF 208
DB 551 FTFPTPLPDGTVVNVVVARDAAGNSPPASVTVDVAPATPTVDPSNGTTL----- 600
QY 209 RSGELENRITVAQHIWSAELPAHWIVPGLNVLVKGSLGSLNDIKIGAPCELLHTID 268
DB 601 -SGTAE-----PGSSVTLTDGN-----GNP----- 619
QY 269 IGMATTPRDRFPAKKAHREYFOTIPVSRVIVNNAAPLHLKVELMPTGELLTD----- 323
DB 620 IQGVTA-----DGSQNWTFSTPLPNCVTVNATAD-----PSGNASSPASVT 664
QY 324 -----ADPGNGGWHSGTMRQRIKGLVSHGIDNANYGLNSTAGIGENSHPHYVAQL 374
DB 665 DAVAPATPVVNPNSGTTLSGT-----AEPGATVTLTDG-RGNP--IQGV 705
QY 375 AAHNSRGNYA-----NGIOVH-----GGSGGGGIVTLDS-----TL 405
DB 706 TADGS-GNNSFTPTPLPNCVTVNATADASNTSAGSSVTVDVAPATPVINPSNGTTL 764
QY 406 GNEFSHEVGHNYGLHYDGFSGVHRSANNNSTWGDGKRKIPNFPYS-QTNEKSC 464
DB 765 SG--TAEPGSSVTL--TDGNGNPICQVTAADGSGNW-----SFTSTPLADGT 808
QY 465 LNNQCEPFDGCHKFGFADAMAGGSPSAANFTWYTPNNSALIQRFFENKAVFDSRSSTGF 524
DB 809 VNATATDP-----AGNTSGOGSTTVDGVAFTPTV-----NLNSGSLSGTA- 850
QY 525 SKWNADTQEMEPYEHTIDRAEQITASVNELSKAELMAEYAVVVKHVMNGNWRTNIYI 584
DB 851 -----EPGS-----TVILTDCNGNPINAEVTAD-----GSGNWT---YT 880
QY 585 PTASADNRGSIILTINHEAGYNSYLFINGDEKVVSOQYKKSFYSDGQFKERDVTDTREAR 644
DB 881 PSTPIANGTVNVVAADAGNS-----SPGASVT-----VDSQAPA 916
QY 645 KPFQFQVPTVTLGVYDPEGTLSYIYPAMYGAFTYSDDSQNLSDNDQCLOQVDTKEGO 704

DB 917 APVNPSTGTTLSGTAEPCATV-----TLTDGNGNFIG---QVTAD-GSGN 958
QY 705 LRFRLANRANNTVMNKHINPTESOPTQATLVCCNKKILDTKSLTPAPEGLTYVNGQA 764
DB 959 WSTPTGCTPLANGTVN-----ATASDPTGNTSAPASTIVD--SVAPAAP-VVNPNGAE 1009
QY 765 LPAKENEGCIVSVNSCKRYCLPVGORSBSYSLPDM 798
DB 1010 ISGTAEPCATVTLTDGSGN--PIGQVTAADGSGNW 1041

RESULI 14
I40805
collagenase - Clostridium histolyticum
C:Species: Clostridium histolyticum
C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 21-Jul-2000
C:Accession: I40805; T44355
R:Yoshihara, K.; Matsushita, O.; Minami, J.; Okabe, A.
J. Bacteriol. 176, 6489-6496, 1994
A:Title: Cloning and nucleotide sequence analysis of the colH gene from Clostridium h
A:Reference number: I40805; MUID:95050206; PMID:7961400
A:Accession: I40805
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1021 <RES>
A:Cross-references: GB:D29981; NID:9563954; PIDN:BAA06251.1; PID:g710023
R:Matsushita, O.; Jung, C.M.; Katayama, S.; Minami, J.; Takahashi, Y.; Okabe, A.
J. Bacteriol. 181, 923-933, 1999
A:Title: Gene duplication and multiplicity of collagenases in Clostridium histolyticu
A:Reference number: 222752; MUID:99121032; PMID:9922257
A:Accession: T44355
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1021 <MAT>
A:Cross-references: EMBL:AB014075; NID:93868863; PIDN:BAA34542.1; PID:g3892646
A:Experimental source: strain JCM 1403
C:Genetics:
A:Gene: colH
C:Superfamily: microbial collagenase

Query Match 3.2%; Score 148; DB 2; Length 1021;
Best Local Similarity 18.6%; Pred. No. 0.22;
Matches 169; Conservative 121; Mismatches 330; Indels 288; Gaps 46;

QY 74 DDNNKILGTLTPSSLPDPTIYHLDGVPEGDDFTPHNGTKKIINTVAEVNKLSDASGS 133
DB 263 DDNNSWI-----IDNGIYHI--APLGLK---HSNNKIGIETLTVMK----- 299
QY 134 SIHSLTNALVEIHTANGRWVRIYLPQGPDLLEGKW---RFVSSAGYS---STVFYGD 187
DB 300 -VYPYLSMQ-----HLOSADQIKRHY--DSKDAEGNKIPLDKFKKGGKCYKPKTYTFDD 351
QY 188 RKVTLVSGNTL-----LFKYNGOMFRSGELENRITVAQHIWSAELPAHWIVPG 237
DB 352 GKVIKAGARVEEKYKRLYASKEVNSQFRR-----VYG 386
QY 238 LNLVIKOGNLSGRNDIKIGAPGELLHTI-----DIGMLTTPRDRFPAKKAHRE 29C
DB 387 IDKPLEGPNDDILTIVYNSPEYKLSVLYGVDYTNMGMYIEPCTF--FTYEREAQES 445
QY 291 YFQTIPTVSRMIVNYY-----APLHLKVELMPTCELLTDMDCGNGHSGHSTHQRIGREL 344
DB 446 TYTLEELFRHEYTHYLOGRYAVFGWGRTKLYDNDRLTWTEG-----GAEL 492
QY 345 VSHGIDNANYGLNSTAGIGENSHPHYVAQIAAHNSRGN--YANGIOVHSGSGGGIVTLTD 402
DB 493 FA-----GSTRTSGLIPRKS--IVSNI--HNTTRNNRYKLSOTVH----- 528
QY 403 STLGNFESHEVGHNYGLHYVD-----GFKGVSFHSRSEANNSTWGDGDKRFPNFI 455
DB 529 SKYGASFEF---YNYAC-MFMDYMYNMDGILNKLNDLAKNNDVD-GYDN-----Y 574

QY 456 PSQTNEKSLNNQCQEPDPGHGFGDAMAGSPFSAANRFTMYT -PNSAIIQRFEENKA 514
Db : ||: | : : | : | : | : | : | : | : | : | :
QY 575 IRLSSNYALNDKYODHMOERIDNTENLT--VPFVADDYLVRHAYKNPNEYSEISVAK 632
Db : ||: | : : | : | : | : | : | : | : | : | :
QY 515 VPDSSSTGFSKWNADTOEMEPYEHTIDRAEQIT--ASVNELSESMAELNAEYAVVKV- 571
Db : ||: | : : | : | : | : | : | : | : | : | :
QY 633 LKDAKSEVKKSQ-----YFSTFTLRGSYTGASKGKLEDKAMNKFIDDSLKKLD 682
Db : ||: | : : | : | : | : | : | : | : | : | :
QY 572 -HMWNGNWTRNIYIPTASADNRGSIILTINHEAGYSYLFINGDEX-----VVSQGYKKS 624
Db : ||: | : : | : | : | : | : | : | : | : | :
QY 683 TYSWSGCYKTLTAFTNYKVDSN---RVTYDVVFHGYPNEGSDSKNSLPYGKINGTYKGT 739
Db : ||: | : : | : | : | : | : | : | : | : | :
QY 625 -----FVSDGQF-----WKERD-----VVD----- 639
Db : ||: | : : | : | : | : | : | : | : | : | :
QY 740 EKEIKFSSEGSDPGDKIVSYEMDFDGCKNSNEENPEHSYDKVGITYTVKLVTDDKGES 799
Db : ||: | : : | : | : | : | : | : | : | : | :
QY 640 -----TREARKPEQFGVPVTTL-----VGYDDPECTLSIYYIPANYGAY 678
Db : ||: | : : | : | : | : | : | : | : | : | :
QY 800 SVSTTTAEIKDLSENKLPVIYMHVPKSGALNKVVYFGKGYTPDGGSIAGIQWD----F 854
Db : ||: | : : | : | : | : | : | : | : | : | :
QY 679 GFTYSDSGLNSDNCOLOVDTKEGQLRFLANHRANNITVMKHFHNVPETESQPTATLV 738
Db : ||: | : : | : | : | : | : | : | : | : | :
QY 855 G-----DGSDFSSEQNPSHVYTKKGEYTVTL-----RVMD-----SSGQMSEKTM- 894
Db : ||: | : : | : | : | : | : | : | : | : | :
QY 739 CNKKILDTKSLTP-----APEGLTYVNGOALPAKENEGCIVSVNGSKRYCLPV---GO- 789
Db : ||: | : : | : | : | : | : | : | : | : | :
QY 895 ---KIKITDPVPICTEKFPNNSKETASGPVPGIPVSGTIENTSDDDYFYFDVITPGEV 951
Db : ||: | : : | : | : | : | : | : | : | : | :
QY 790 -----RSGYSLPMWYVGQEYVDVSDGAKAKVLISDMNLNFRIGEFVGVGNVPADMKKVKA 844
Db : ||: | : : | : | : | : | : | : | : | : | :
QY 952 KIDINKLYGGATWV-----YDENNNNAVSYATDDGQNLGSL---GKFKADKPGRYYIHLYM 1003
Db : ||: | : : | : | : | : | : | : | : | : | :
QY 845 WNGGYLDF 852
Db : ||: | : : | : | : | : | : | : | : | : | :
QY 1004 FNGSYHPY 1011
Db : ||: | : : | : | : | : | : | : | : | : | :

RESULT 15
AC1533
surface protein (LPXTG motif) [imported] - Listeria innocua (strain Clip11262)
C/Species: Listeria innocua
C/Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C/Accession: AC1533
R:/Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker,
J.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:/Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma-
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Valquez-Boland, J.A.; Voss, H.; Wehlend,
B.; Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker,
J.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
A:/Title: Comparative genomics of Listeria species.
A:/Reference number: AB1077; MUID:21537279; PMID:11679669
A:/Accession: AC1533
A:/Status: preliminary
A:/Molecule type: DNA
A:/Residues: 1-1386 <GLA>
A:/Cross-references: GB:AL592022; PIDN:CAC96035.1; PID:g16413254; GSPDB:GN00178
A:/Experimental source: strain Clip11262
C:/Genetics:
A:/Gene: lin0803

Query Match	3.28;	Score 147;	DB 2;	Length 1386;
Best Local Similarity	18.9%;	Pred. No. 0.4;		
Matches 166;	Conservative 113;	Mismatches 297;	Indels 300;	Gaps 44;

QY	110	PHNGTKKLIINTVAEWNKLSA-----SGSSI-----HSHLTNNALVEIHT--ANGRWVPDI	158
		: :	: :
DB	406	PYNGTWTNPTRKRTI--LSDSLTKYDGMTGYHWSAVNNIIAIDSTIYSGDDWKAED	463
		: :	: :
QY	159	YLPQPGDLEGKWRVF--VSSAGSYSTVFYGRKVTLSVGNLTLLFKVYVNGQWFRSGLENN	216
		: :	: :
DB	464	NFISATDKDFNLIDPLVTVTGSDVDTTPGEYETISVNGLTTITTVTKENQASVVAEN	523
		: :	: :
QY	217	RITYAQHIWSAELPAHIVFGLNALVIKQGLSLRLNDIK-----IGAPGEL--LLHT	266

Search completed: October 8, 2003, 16:28:19
Job time : 25.2134 secs

Search completed: Octob
Job time : 25.2134 secs

Db	524	STIYTKESWAE-----DNFVSATNKGKGIADISSVTVTGEVDVNTPGDYEIYNT	573
Qy	267	ID-----IGMLTPDRDFDPAKDEAREHREYFOTIIPVSRMI	301
Db	574	IDGVSTKIIVTVKEDSSIEAKDSILYIGDTWNSKDNFISATDKDGNPVDKDKIVBG-T	632
Qy	302	VNNYAP-----LHLKEVMLPTGELLTDMDPGNGGWHSGT	335
Db	633	VNTIKPGTNKVTVLYGNOSKEVITIVKADQSTLEAKDSIIYTG-----DKWNAKDNFIT	686
Qy	336	MRORIG-----KELVSHG-IDNANYCLNSTAGLGENSHPHYVVAQLAAH-----	377
Db	687	AIDKDGMPVDKIDIEVEGTDTTKPGTNKITIYIGNLSKEVITIVKANOATLEAKDSALY	746
Qy	378	-----NSRGNRYANG-----IOYHGGSGGGVITLDST-----LGNEFS	410
Db	747	VGDTWNSKDNFISATDKDGPVDFDKIVEG-----TVDTTKAGTNKVTVLYGNQ-S	797
Qy	411	HEVGHNTGLGHYDVGFGSVHRSBAENNSTGWGDGKRFIPNFPYSPOTNEKSLNNQOC	470
Db	798	KEVTNVYKADQTTILEVKDSYI-----YTGDWKAAEDNFI-SATDK-----	836
Qy	471	EPFDGKFGF-DAMAGG---SPFSAANRET-MYTPNSSAIQORFFENKAVFDRSS---T	522
Db	837	---TGNSVGFKDVKVEGTVDTTKAGTNKVTVLYYANOSKEVITIVKADQATLEAKDSIIYT	893
Qy	523	GFSKWNA-----DQOMEPEYEHT-----IDRABOQITASV	551
Db	894	G-DKWKAAEDNFISATDKAGKTIDFNKIKVEGTDITTKAGNYDITYSYSGVTRSTELSKTI	952
Qy	552	NELSESKMAELMAEYAVVYHMGNGN-W-TRNIYIPTASADN-----RGSILITIN	599
Db	953	TVTVKKQVNLKADST-----LYEGDKRWAKDNFVSATDKOGNTVDYKAIKVEGTVWTK	1008
Qy	600	-----HEAGYNSYLFIN---GDEKWSQGYKKSFSVDGQFWKERDVDPVTRARKPEQ	648
Db	1009	AGTYKITIYSAGISKTIITVTLANQTKIVA---KDLTIYEGDNWNEQD---NFWSATDK	1061
Qy	649	EG-----VPVTVLYGYVDPEGTVLSYIYPAMYGAYGFTYS-----DDSQN	688
Db	1062	FQQAIDFNSKVTGSDVDIOTP-----GKIRITYSIEGASTIITVTLAQSN	1108
Qy	689	LSDNDQOLQVDTK-EGQLRTRFLANHRANNTVMNKFHINVPESQPTQATLCNNKILDTK	747
Db	1109	LVANKNSTIYGDKKQSKDNFVSATDK-----YGRKIDISLLTVTGTVD-T	1152
Qy	748	SUTPAPEGLTYTVNGQ-----ALPAKRENGECIVSVNS	779
Db	1153	-TTPEYEITYSVNGLTITITVTVKNOANIVAEADS	1187

Search completed: October 8, 2003, 16:28:19
Job time : 25.2134 secs